НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО

Факультет інформатики та обчислювальної техніки
(назва факультету, інституту)

Кафедра автоматизованих систем обробки інформації і управління
(назва кафедри)

"На правах рукопису"
УДК *519.854*

«До захисту допущено»
Завідувач кафедри

_____ О.А.Павлов
(підпис) (ініціали, прізвище)

“ ” _____ 20 *18* р.

МАГІСТЕРСЬКА ДИСЕРТАЦІЯ

на здобуття ступеня магістра

за спеціальністю *122 Комп'ютерні науки та інформаційні технології*
(код та назва спеціальності)

спеціалізацією *Інформаційні управляючі системи та технології*
(код та назва спеціалізації)

на тему: *Задача побудови багатовимірної поліноміальної регресії по надлишковому опису*

Виконав: студент *VI* курсу групи ІС-63м
(шифр групи)

Коваленко Дмитро Андрійович _____
(прізвище, ім'я, по батькові) (підпис)

Науковий керівник *проф., д.т.н., проф. Павлов О. А.* _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Консультант *к.т.н., доц. Жданова О.Г.* _____
(науковий ступінь, вчене звання, прізвище, ініціали) (підпис)

Рецензент _____ _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.
Студент _____
(підпис)

Київ – 2018

РЕФЕРАТ

Магістерська дисертація: ХХ с., ХХ рис., ХХ табл., ХХ додатків, ХХ джерел.

Актуальність. Проблема знаходження істинної закономірності за результатами експериментів є універсальною. Немає ні однієї області діяльності людини, в якій так чи інакше не виникала б ця задача. В економічних, соціологічних та природничих науках часто вирішують задачу виявлення чинників, що визначають рівень і динаміку процесів. Таке завдання найчастіше вирішується методами кореляційного, регресійного, факторного і компонентного аналізу. Завдання регресійного аналізу полягає в побудові моделі, що дозволяє за значеннями незалежних показників отримувати оцінки значень залежної змінної. Різні аспекти розв'язку цієї проблеми розглядаються в таких науках, як математична статистика, теорія управління, теорія штучного інтелекту. В рамках теорії імовірності ця задача формулюється як оцінка лінії регресії по результатам статистичних експериментів і на практиці є областю прикладного регресійного аналізу.

Проблема відтворення невідомої залежності формулюється як класична задача прикладного регресійного аналізу: відтворення багатовимірної поліноміальної регресії по надлишковому опису і з довільно розподіленою похибкою. По результатам активних експериментів необхідно знайти невідомі коефіцієнти, частина з яких тотожно дорівнює нулю і невідома досліднику. На відміну від кореляційного аналізу не з'ясовує чи істотний зв'язок, а займається пошуком моделі цього зв'язку, вираженої у функції регресії. Регресійний аналіз використовується в тому випадку, якщо відношення між змінними можуть бути виражені кількісно у виді деякої комбінації цих змінних. Отримана комбінація використовується для передбачення значення, що може приймати цільова (залежна) змінна, яка обчислюється на заданому наборі значень вхідних (незалежних) змінних. У найпростішому випадку для цього використовуються стандартні статистичні методи, такі як лінійна регресія. На жаль, більшість реальних моделей не вкладаються в рамки лінійної регресії. Наприклад, розміри продажів чи фондові ціни дуже складні для передбачення, оскільки можуть

залежати від комплексу взаємозв'язків множин змінних. Таким чином, необхідні комплексні методи для передбачення майбутніх значень.

Саме тому розробка алгоритмів, які б допомогли вирішити цю проблему - проблему регресії багатьох змінних - є дуже актуальною у наш час і залишатиметься такою ще довго. Задача, яка постає перед нами, є дуже складною, адже загальний опис ситуації, який було зазначено вище, не показує усіх можливих складностей відтворення реальної залежностей складних процесів. Саме тому і досі не існує алгоритму, який міг би знайти дуже гарний розв’язок для усіх формулювань цієї задачі за прийнятний час. Проте вже було проведено дослідження та розроблено деякі можливі алгоритми розв’язання даної задачі, які для деяких варіації дають дуже гарні результати. Необхідно продовжувати дослідження та намагатися покращувати вже отримані результати.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «СКЛАДНОВИРІШУВАНІ ЗАДАЧІ КОМБІНАТОРНОЇ ОПТИМІЗАЦІЇ В ПОАНУВАННІ І ПРИЙНЯТТІ РІШЕНЬ» (№ 143).

Мета дослідження – підвищення ефективності алгоритму багатовимірної поліменіальної регресії.

Для досягнення мети необхідно виконати наступні **завдання:**

- виконати огляд відомих результатів з поставленої задачі;
- формалізувати задачу відновлення залежності за даними експериментів;
- розробити алгоритм для розв’язання поставленої задачі що базується на активному експерименті;
- розробити програмну реалізацію алгоритмів та моделей у вигляді, що може використовуватися в різних галузях для розв’язання даної задачі;

- провести порівняльний аналіз існуючих алгоримів та нового алгоритму;
- виконати аналіз отриманих результатів.

Об'єкт дослідження – процес відновлення реальної залежності складних процесів за даними експериментів.

Предмет дослідження – складні економічні або природничі процеси.

Методи дослідження – стохастичне програмування, порівняльний аналіз, статистичний аналіз.

Наукова новизна одержаних результатів полягає у створенні та оптимізації принципово нового підходу до відтворення складних природних та економічних процесів у вигляді лінії регресії. Сучасні рішення даної проблеми не є евристичними алгоритмами та не є статистично коректними.

РЕГРЕСІЯ, БАГАТОВИМІРНА ПОЛІНОМІАЛЬНА РЕГРЕСІЯ, ЕКСПЕРИМЕНТ, АКТИВНИЙ ЕКСПЕРИМЕНТ, ПОЛІНОМИ ФОРСАЙТА, ПОВТОРЮВАНІ ЕКСПЕРИМЕНТИ, РЕКУРЕНТНІ СПІВВІДНОШЕННЯ

ABSTRACT

Актуальність. Проблема знаходження істинної закономірності за результатами експериментів є універсальною. Немає ні однієї області діяльності людини, в якій так чи інакше не виникала б ця задача. В економічних, соціологічних та природничих науках часто вирішують задачу виявлення чинників, що визначають рівень і динаміку процесів. Таке завдання найчастіше вирішується методами кореляційного, регресійного, факторного і компонентного аналізу. Завдання регресійного аналізу полягає в побудові моделі, що дозволяє за значеннями незалежних показників отримувати оцінки значень залежної змінної. Різні аспекти розв'язку цієї проблеми розглядаються в таких науках, як математична статистика, теорія управління, теорія штучного інтелекту. В рамках теорії імовірності ця задача формулюється як оцінка лінії

регресії по результатам статистичних експериментів і на практиці є областю прикладного регресійного аналізу.

Проблема відтворення невідомої залежності формулюється як класична задача прикладного регресійного аналізу: відтворення багатовимірної поліноміальної регресії по надлишковому опису і з довільно розподіленою похибкою. По результатам активних експериментів необхідно знайти невідомі коефіцієнти, частина з яких тотожно дорівнює нулю і невідома досліднику. На відміну від кореляційного аналізу не з'ясовує чи істотний зв'язок, а займається пошуком моделі цього зв'язку, вираженої у функції регресії. Регресійний аналіз використовується в тому випадку, якщо відношення між змінними можуть бути виражені кількісно у виді деякої комбінації цих змінних. Отримана комбінація використовується для передбачення значення, що може приймати цільова (залежна) змінна, яка обчислюється на заданому наборі значень вхідних (незалежних) змінних. У найпростішому випадку для цього використовуються стандартні статистичні методи, такі як лінійна регресія. На жаль, більшість реальних моделей не вкладаються в рамки лінійної регресії. Наприклад, розміри продажів чи фондові ціни дуже складні для передбачення, оскільки можуть залежати від комплексу взаємозв'язків множин змінних. Таким чином, необхідні комплексні методи для передбачення майбутніх значень.

Саме тому розробка алгоритмів, які б допомогли вирішити цю проблему - проблему регресії багатьох змінних - є дуже актуальною у наш час і залишатиметься такою ще довго. Задача, яка постає перед нами, є дуже складною, адже загальний опис ситуації, який було зазначено вище, не показує усіх можливих складностей відтворення реальної залежностей складних процесів. Саме тому і досі не існує алгоритму, який міг би знайти дуже гарний розв’язок для усіх формулювань цієї задачі за прийнятний час. Проте вже було проведено дослідження та розроблено деякі можливі алгоритми розв’язання даної задачі, які для деяких варіації дають дуже гарні результати. Необхідно

продовжувати дослідження та намагатися покращувати вже отримані результати.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «СКЛАДНОВИРІШУВАНІ ЗАДАЧІ КОМБІНАТОРНОЇ ОПТИМІЗАЦІЇ В ПОАНУВАННІ І ПРИЙНЯТТІ РІШЕНЬ» (№ 143).

Мета дослідження – підвищення ефективності алгоритму багатовимірної поліменіальної регресії.

Для досягнення мети необхідно виконати наступні **завдання:**

- виконати огляд відомих результатів з поставленої задачі;
- формалізувати задачу відновлення залежності за даними експериментів;
- розробити алгоритм для розв’язання поставленої задачі що базується на активному експерименті;
- розробити програмну реалізацію алгоритмів та моделей у вигляді, що може використовуватися в різних галузях для розв’язання даної задачі;
- провести порівняльний аналіз існуючих алгоримів та нового алгоритму;
- виконати аналіз отриманих результатів.

Об’єкт дослідження – процес відновлення реальної залежності складних процесів за даними експериментів.

Предмет дослідження – складні економічні або природничі процеси.

Методи дослідження – стохастичне програмування, порівняльний аналіз, статистичний аналіз.

Наукова новизна одержаних результатів полягає у створенні та оптимізації принципово нового підходу до відтворення складних природних та

економічних процесів у вигляді лінії регресії. Сучасні рішення даної проблеми не є евристичними алгоритмами та не є статистично коректними.

РЕГРЕСІЯ, БАГАТОВИМІРНА ПОЛІНОМІАЛЬНА РЕГРЕСІЯ, ЕКСПЕРИМЕНТ, АКТИВНИЙ ЕКСПЕРИМЕНТ, ПОЛІНОМИ ФОРСАЙТА, ПОВТОРЮВАНІ ЕКСПЕРИМЕНТИ, РЕКУРЕНТНІ СПІВВІДНОШЕННЯ

ЗМІСТ

ВСТУП

1 ОГЛЯД ЛІТЕРАТУРИ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Задача відновлення залежності за даними експерименту для декількох змінних є давно відомою та важливою задачею. У найпростішому випадку для цього використовуються стандартні статистичні методи, такі як лінійна регресія. На жаль, більшість реальних моделей не вкладаються в рамки лінійної регресії. Наприклад, розміри продажів чи фондові ціни дуже складні для передбачення, оскільки можуть залежати від комплексу взаємозв'язків множин змінних. Таким чином, необхідні комплексні методи для передбачення майбутніх значень. Ця задача ускладнюється експоненційно в залежності від кількості змінних, тому її вирішення для складних систем з великою кількістю змінних є практично неможливим. Для її розв’язання було розроблено безліч алгоритмів – більшість алгоритмів є вдалими евристиками для конкретних задач та галузей. Наведемо деякі попередні дослідження та методи її розв’язання.

Поняття кореляції в прийнятому нами значенні з'явилося майже в середині XIX століття завдяки роботам сера Френсіса Гальтона (двоюрідного брата Чарльза Дарвіна) і Карла Пірсона. Ф. Гальтон застосував для кореляції наступну форму запису: co-relation, звідки стає зрозумілим значення цього виразу - зв'язок, співвідношення. Спочатку дослідження кореляції проводились в галузі природничих наук, перш за все в біології. Лише пізніше застосування методів кореляційного аналізу поширилося на економіку, де вони привели до вельми корисним результатами.

Поняття регресії також сходить до Ф. Гальтону. Після знайомства з книгою Чарльза Дарвіна «Походження видів» [1] в 1859 р Ф. Гальтона стала займати думка про те, чому люди з покоління в покоління не сильно розрізняються за зовнішнім виглядом і природним здібностям. Це привело його до вивчення спадковості [2]. Зокрема, він зайнявся з'ясуванням залежності зростання дітей від зростання батьків. За логікою діти повинні бути кожен раз дуже схожі на своїх батьків. Високі батьки повинні мати високих дітей, а низькорослі батьки - дітей низького зросту. При такому стані речей через кілька поколінь ми мали б, з

одного боку, рід велетнів, а з іншого - рід карликів. Але незабаром в результаті великих статистичних досліджень і дослідів над тваринами Ф. Гальтон переконався, що такої тенденції немає, а, скоріше, навпаки, діти дуже високих або дуже низьких батьків в середньому мають менш високий або відповідно менше низький зріст. Крім того, ухилення зростання дітей не таким значним, як ухилення зростання їх батьків від середнього зросту досліджених осіб. Це рух назад в напрямку до середнього Ф. Гальтон назвав регресією (to regress - рухатися в зворотному напрямку) [3], [4].

Лінійна регресія [5] є дуже популярною та вивченою моделлю регресії. В багатьох галузях науки нам необхідно дослідити як зміни у одній змінній відображаються на іншій змінній. Іноді дві змінні звязані точно прямолінійною залежністю, Наприклад, якщо опір кола підтримується на сталому рівні, сила струму пропорційна напрузі [6], I=V/R. Якщо нам невідомий закон Ома, ми можемо отримати його емпірично, змінюючи напругу та спостерігаючи за силою струму, не змінюючи при цьому опір, ми можемо спостерігати що графік сили струму в залежності від напруги більш-менш приймає форму прямої лінії через точку (0,0). «Більш-менш» використовується для тому що при проведені вимірів виникають невеликі похибки.

Задача лінійної регресії у такому вигляді є найбільш поширеною та вивченою (особливо в економетриці). Вивчені властивості параметрів, що отримуються різними методами при припущеннях про ймовірносні характеристики факторів та випадкових похибок моделі.

Найвідоміший метод оцінки параметрів лінійної регерсії є *метод найменших квадратів* [7]. Перша робота, в якій був використаний метод найменшіх квадратів належить Лежандру. В 1805р. в статті “Нові методи визначення орбіт комет” [8] він писав “Після того як повністю використані умови задачі, необхідно визначити коефіцієнти так, щоб виличини їх помилок були найменшімі із можливих. Для цього нами вказаний простий спосіб, який полягає в знаходженні мінімума сумми квадратів помилок”.

Першим, хто намагався поставити даний метод на міцну математичну основу був, судячи із всього, Р. Адрейн. В його роботі “Дослідження, що стосуються ймовірностей помилок, які зє при змінах” (1808р.) [9] можна знайти грубий “доказ” того, що в деяких умовах ці помилки е підпорядковуются розподілу:

$\varepsilon = C - h2\varepsilon 2,$

тобто нормальному розподілу. Звідси автоматично витікав метод найменших квадратів, який давав “найбільш ймовірні”, або найбільш правдоподібні рішення для невідомих параметрів лінійної форми.

В 1809р. Гаусс в відомій роботі з обчислення орбіт дав друге обгрунтування закону розподілення помилок. Крім того, Гаусс відстоював використання методу найменших квадратів з 1795р. [10]

В 1792р. Лаплас застосовував критерій суми абсолютних значень а не суми квадратів [11]. В 1831р. Коші запропонував метод, в якому також використовувалась максимізація суми помилок без урахування знаку – цей метод іноді використовують і тепер. [12]

Подалі використання методу найменшіх квадратів звязане з іменем Лапласа, який в 1812р. в своїй роботі “Аналітична теорія ймовірностей” показав, що цей метод дозволяє знайти незміщені оцінки незважаючи на тип вхідного розподілу.

Гаусс опублікував свої міркування щодо даного методу в 1821р. Не використовуючи такі поняння як дисперсія та не прибігаючи до матричної алгебри, він довів, що серед класу оцінок, які являються:

a) лінійними комбінаціями вхідних даних;
b) незміщеними оцінками параметрів;

оцінки, отримані методом найменших квадратів, мають найменші похибки. Найбільш важлива характеристика оцінок, що отримані методом найменших квадратів, полягає у незалежності від типу розподілу.

В більш загальному вигляді теорема була доведена в 1912р. А. Марковим і в даний час відома як теорема Гауса-Маркова [13]. Ця теорема є центральною в методі найменших квадратів, наведемо її: Теорема Гауса-Маркова [14]:

Якщо виконуються дані передумови:

1. Математичне сподівання випадкової похибки e_i дорівнює нулю: $M(e_i) = 0$ для всіх спостережень.
2. Дисперсія випадкової похибки е постійна: $D(\varepsilon i) = D(\varepsilon j) = \sigma^2 = \text{const}$ для будь-яких спостережень i-j. Умова незалежності помилок від порядкового номера спостереження називається гомоскедатичністю
3. Випадкові відхилення εі і εj є назележними один від одного для $i \neq j$. Виконання цієї передумови передбачає, що відсутній систематичний звязок між будь-якими випадковими похибками. Велечина та знак відхилення не повинні бути причиною величини чи знака іншого відхилення.
4. Випадкова похибка повинна бути незалежна від пояснювальних (незалежних) змінних. Зазвичай ця умова виконується якщо незалежні змінні не є випадковими величинами.
5. Модель являється лінійною відносно параметрів. Для випадку багатовимірної регресії накладаються ще дві умови:
6. Відсутність мультиколінеарності. Між пояснювальними (незалежними) змінними не повинна бути сильна лінійна залежність.
7. Випадкові відхилення εі, $i = 1, 2, \ldots, n$, розподілені нормально. Виконання данної умови важлива для перевірки статистичних гіпотез та побудови інтервальних оцінок.

Поряд з виконанням вказаних передумов при побудові класичних лінійних регресійних моделей робляться такі припущення:

- Пояснювальні (незалежні) змінні не є випадковими величинами;

- Число спостережень (розмір вибірки) набагато більше числа незалежних змінних (числа факторів рівняння);
- Відсутні помилки специфікації, тобто правильно вибраний вид рівняння і в нього включені все необхідні змінні;
- Часто припускають що спостережень повинно бути в 5-6 разів більше ніж кількість параметрів рівняння.

Якщо виконуються вказані умови, то оцінки, отримані методом найменших квадратів мають такі властивості:

1. Оцінки є незміщеними, тобто $M(b_1) = \beta_1$, $M(b_0) = \beta_0$ (математичне очікування оцінок параметрів рівні їх теоретичним значенням). Це витікає з того, що $M(\varepsilon_i) = 0$, і показує що систематична помилка в лінії регресії відсутня.
2. Оцінка переметрів
3. Дисперсія оцінок параметрів при возрастанні числа n спостережень прямує до нуля $D(b_0) \to 0$, $D(b_1) \to 0$ при $n \to \infty$. Інакше кажучи, при збільшенні обєма виборки надійність оцінок покращуються.
4. Оцінки параметрів ефективні, тобто мають найменшу дисперсію по відношенню до інших оцінок даних параметрів, лінійних відносно y_i.

Великий вклад в розвиток методу був здійснений в 1934р. Ейткеном, який узагальнив теорему на випадок корельованих результатів спостережень з різними дисперсіями. Робота М. Мерримана [15] збирає історичний розвиток методу найменших квадратів та дає критичні зауваження.

В сучасних роботах [16], [17] статистики та прикладного регресійного аналізу, метод найменших квадратів є базисом регресійного аналізу. Зокрема, в [17] приведений повний алгоритм методу найменших квадратів та описані методи підвищення точності та спрощення обчислень даного методу.

Ортогональні поліноми

В [16], [18], [19] був наведений метод значного покращення результатів регресійного аналізу за допомогою ортогональних поліномів.

При оцінці коефіцієнтів поліноміальної регресії та використанні степенів незалежної змінної як вхідних даних МНК, матриця (1) є погано обумовленою, часто rang(A) < m при великих r, більш того, погана обумовленість може погіршитися зі зростанням обєму вибірки. Це призводить до великих похибок при знаходженні коефіцієнтів методом найменших квадратів.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & \dots & x_2^m \\ 1 & x_3 & x_3^2 & \dots & x_3^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

Вирішення даної проблеми було запропоновано D. Van Der Reyden у його роботі “Curve Fitting by the Orthogonal Polynoms of Least Squares” [20]. У його роботі за допомогою методу Грама-Шмідта [21] модель зводиться до:

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \alpha_2 P_2(x_i) + \dots + \alpha_k P_k(x_i) + \varepsilon_i,\ i = 1, 2, \dots, n$$

Де $P_u(x_i)$ - u-тий ортогональний поліном, що визначається як:

$$\sum_{i=1}^{n} P_r(x_i) P_s(x_i) = 0,\ r \neq s, r, s = 0, 1, 2, \dots, k$$

$$P_0(x_i) = 1.$$

Матриця X приймає вигляд:

$$X = \begin{bmatrix} P_0(x_1) & P_1(x_1) & \cdots & P_k(x_1) \\ P_0(x_2) & P_1(x_2) & \cdots & P_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(x_n) & P_1(x_n) & \cdots & P_k(x_n) \end{bmatrix}.$$

Тоді:

$$X'X = \begin{bmatrix} \sum_{i=1}^{n} P_0^2(x_i) & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^{n} P_1^2(x_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{i=1}^{n} P_k^2(x_i) \end{bmatrix}.$$

Індивідуальні оцінки коефіцієнтів:

$$\hat{\alpha}_j = \frac{\sum_{i=1}^{n} P_j(x_i) y_i}{\sum_{i=1}^{n} P_j^2(x_i)}, \quad j = 0,1,2,...,k$$

Використання даного методу призводить до менших похибок округлення, хоча два методи і еквівалентні алгебраїчно. До того ж, використання ортагональних поліномів дозволяє легко визначити ступінь r апроксимуючого полінома. Загалом, використання ортогональних поліномів приводить до надзвичайно ясному статистичному аналізу.

Рекурентні співвідношення Форсайта

Форсайт запропонував [22], [23] рекурентний метод знаходження ортонональних поліномів,

$$\lambda\,\theta_j(x)=x\,\theta_{j-1}(x)-\alpha\,\theta_{j-1}(x)-\beta\,\theta_{j-2}(x)$$

$$\alpha=\sum_{i=1}^{n}x_i\,\theta_{j-1}^2(x_i);\;\;\beta=\sum_{i=1}^{n}x_i\,\theta_{j-1}(x_i)\theta_{j-2}(x_i);$$

де λ знаходиться:

$$\lambda=\sqrt{\sum_{i-1}^{n}\left(x_i\theta_{j-1}(x_i)-\alpha\theta_{j-1}(x_i)-\beta\theta_{j-2}(x_i)\right)^2}.$$

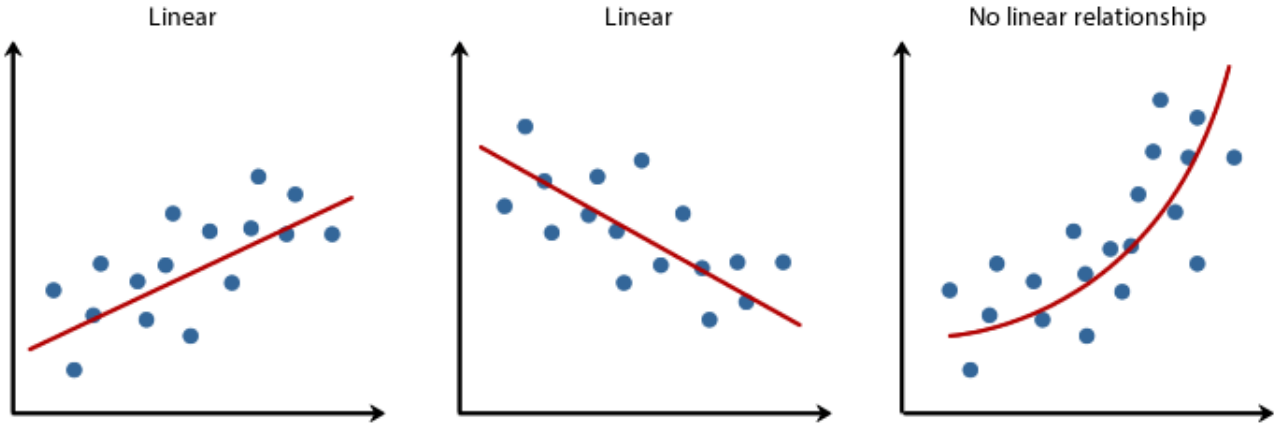

Рис 1.1 - Приклад лінійної та нелінійної регерсії

Найбільш доступний та зрозумілий виклад метод найсменших квадратів був зроблений у книзі Уізборна [24]. Шикокий огляд регресійного аналізу також міститься в книзі Девіса [25], стр 150. Однією з типових монографій на цю тему являється книга Плекета [26]. Виклад охоплює випадок, коли помилки приймають різні ваги і, крім того, коррельовані. Детально розглянуті приклади застосування регресійного аналізу у другому томі книги Кендалла та Стюарта [27]

В більш сучасних задачах стає необхідною апроксимувати результати нелінійної форми. При цьому зазвичай використовується лінеаризація за допомогою розкладу в ряд Тейлора. Регресійні рівняння в такому випадку

вирішуються методом ітерації, описаним в книзі Вільямса [28]. Там же приведені цікаві приклади із області лісоводства.

В ряді задач параметри звязані між собою деяким рівнянням, так що число навідомих параметрів можна зменшити на одиницю. Таке рівняння також може бути нелінійним відносно параметрів. Детально це питання досліджено в главі 3 Жоно і Морелля [29].

Якщо корреляція між помилками являється фукнцією порядку, в якому були отримані експериментальні дані, та такі дані називаються «часовим рядом». Сучасному викладу даного питання присвячена книга Гренандера і Розенблатта [30].

Методи регуляризації

Заслуговують на увагу і методи регуляризації [31], [32], [33]. Регулярізація, в математиці і статистиці, а також в задачах машинного навчання і обернених задачах, означає додавання деякої додаткової інформації, щоб знайти рішення некоректно сформованної задачі, або щоб уникнути перенавчання.

Загалом регулярізуючий параметр *R(f)* додається до звичайної функції втрат:

$$\min_f \sum_{i=1}^{n} V(f(\hat{x}_i), \hat{y}_i) + \lambda R(f)$$

де *V* - функція, що визначає похибку передбачення *f(x)* для значень *y*, (наприклад, квадрати похибок), а параметр λ визначає величину регулярізації. *R(f)* є, зазвичай, штрафом за складність функції *f*, такі як обмеження на гладкість, або на норму векторного простору [34].

Фактично, процедура регулярізаціїдає змогу застосувати лезо Оккама до рішення. З баєсіанської точки зору, багато технік регулярізації є накладанням обмежень на апріорний вигляд розподілу параметрів моделі.

Метод локальної регресії

Методи LOESS та LOWESS (locally weighted scatterplot smoothing) [35], [36] – це два схожих непараметричних методи регресії які поєднують у собі

декілька регресійних моделей у мета-моделі К найближчих сусідів. LOESS це узагальнений метод LOWESS. Обидва методи будуються на класичних методах, такі як лінійний та нелінійний метод найменших квадратів. Вони адресують проблему коли клачисні методи не можуть адекватно оцінити дані. LOESS обєдную простоту лінійного методу найменших квадратів та гнучкість нелінійної регресії. Це стає можливим завдяки застосуванню простих регерсійних моделей на підмножинах вибірки для побудови функції що пояснює детерміновану частину даних точка за точкою. Краса цбого методу полягає у тому що аналітику не потрібно задавати глобальну функцію якої-небудь форми – лише задавати прості моделі для сегметів даних.

Недоліком даних моделей є підвищена кількість обчислень. Через інтенсивність обчислень, даний метод було практично неможливо використовувати у еру розробки алгоритмів найменшіх квадратів. Більшість сучасних регресійних алгоритмів схожі на LOESS в цьому розуминни – нові алгоритми використовують потужності сучасних компютерів для отримання кращих резульатів ніж можуть дати класичні методи.

Метод MARS

Метод адаптивних регресійних сплайнів це форма регерсійного аналізу що була запропонована Жеромом Фрідманом в 1991р. [37], [38]. Це непараметризований метод регресії і може роглядатися як рохширення лінійних моделей що автоматично моделює нелінійні відношення та взаємодію між змінними. При цьому MARS розбиває вибірку на під-множити і використовує методу лінійного регерсійного аналізу.

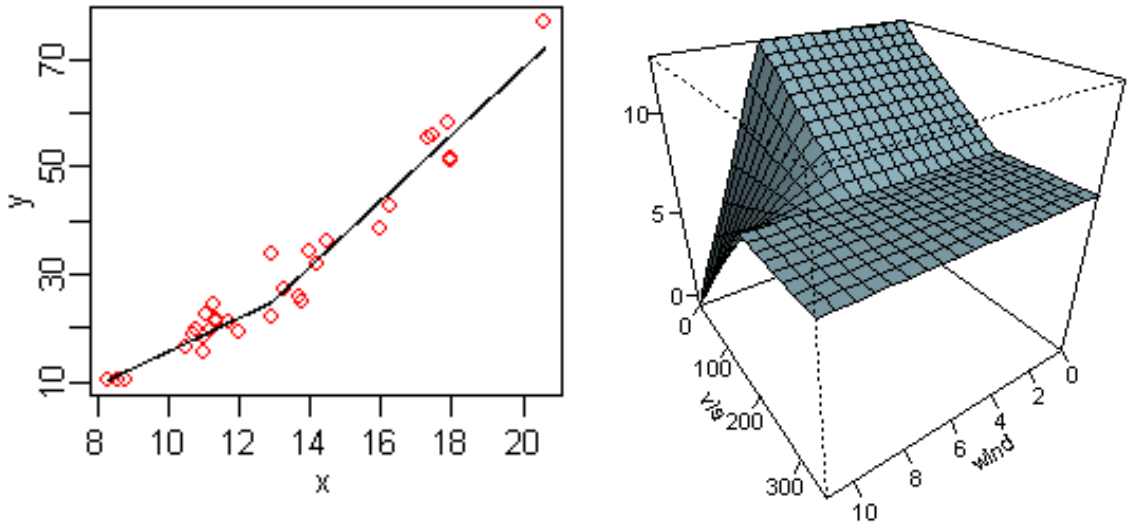

Рис 1.2 – Поліном регресії за методом MARS

Багатовимірна поліноміальна регресія

У [39], [40] пропонується і обґрунтовується ефективний алгоритм вирішення задачі багатовимірного регресійного аналізу, що базується на використанні нормованих поліномів Форсайта [41] в рамках класичного методу найменших квадратів. Основна прикладна особливість методу, запропонованого нижче – відсутність обчислювальної складності, формування вимог до експерименту (плану експерименту), виконання яких призводить до майже точного відтворення закономірності (багатовимірної поліноміальної регресії) при адитивній похибці з довільною скінченною похибкою.

В [42] приведений метод побудови багатовимірної поліноміальної регресії по надлишковому опису в умовах активного експерименту. Алгоритм базується на зведенні багатовимірної регресії до одновимірної за допомогою фіксації всіх змінних окрім однієї. Цей метод має той недолік, що в одновимірних регресіях, що будуються в процесі роботи алгоритму, міститься велика кількість членів з малими значеннями ступенів при *x*. Як буде показано нижче, коефіцієнти таких поліномів погано відтворюються.

В даній роботі пропонується модифікований алгоритм побудови багатовимірної поліноміальної регресії, що базується на ідеях, викладених в [43].

Основна ідея модифікованого методу полягає в фіксації невеликої кількості змінних та одночасній зміні інших (не фіксованих) параметрів. Це призводить до одновимірних регресій з вищими степенями та більш точними оцінками невідомих коефіцієнтів.

Методи машинного навчання для вирішення задачі регресії

Крім методів регерсійного аналізу для вирішення задачі регресії використовуются і методи машинного навчання. В принципі цю задачу може виконати будь-який алгоритм навчання з учителем. Наприклад, в [30], [44] надається приклад використання персептрона для вирішення задачі регресії. Хоча багато-шаровий персептрон і вирішую задачу класифікації, його все ж можна використовувати для задач регресії не використовуючи при цьому активуючу функцію на останньому шарі – вихідні значення при цьому неперервні.

Дерева прийняття рішень можуть бути використані для задачі регресії. При цьому листки дерева приймають неперервні значення. В такому випадку такі дерева називають регерсійними.

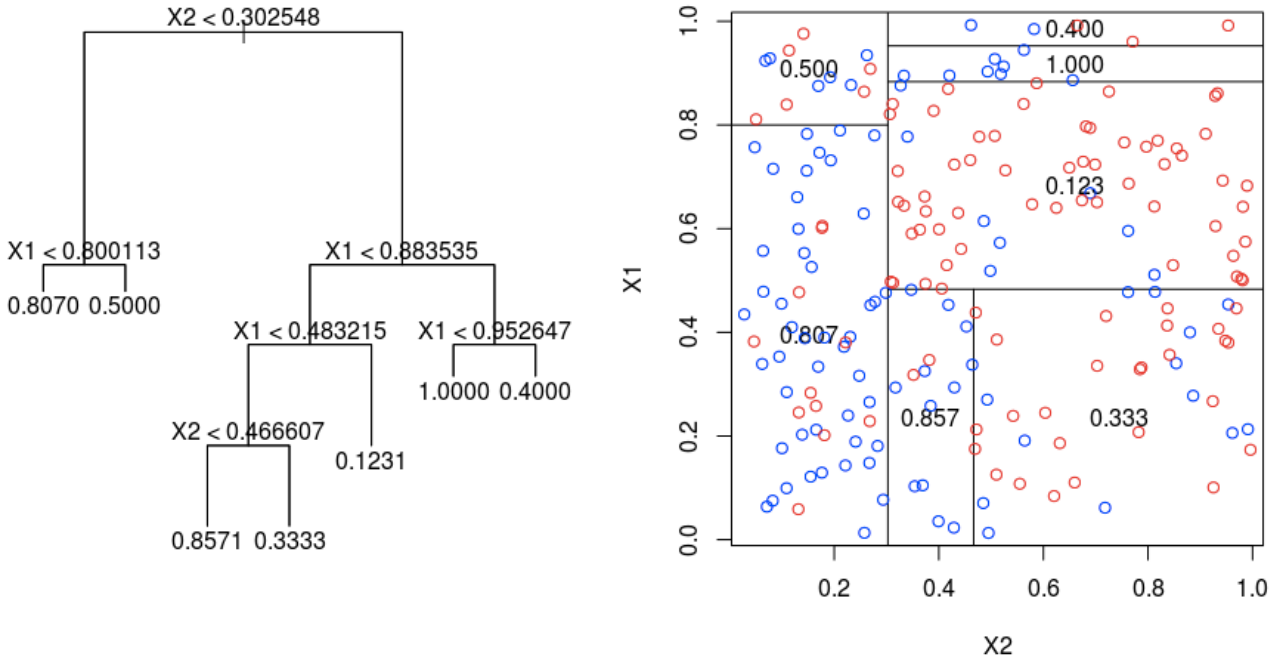

Рис 1.3 – Розбиття множини вибірки за допомогою регресійних дерев

У [45] Aleksandar Peckov описує алгоритм регресії що обєднує MARS та методи машинного навчання, додає декілька вдалих евристик, і досягає при цьому гарних результатів.

МГУА

Метод групового урахування аргументів (МГУА) розробляється академіком НАНУ О.Г.Івахненком та його школою, починаючи з 60-х років[46]. Це типовий метод індуктивного моделювання і один з найбільш ефективних методів структурно-параметричної ідентифікації складних об'єктів, процесів і систем за даними спостережень в умовах неповноти інформації.

В цілому задача ідентифікації полягає у формуванні за даними вибірки деякої множини моделей різної структури

$$\hat{y}_f = f(X, \hat{\theta}_f)$$

і пошукові оптимальної моделі за умовою

$$f^* = \arg \min_{f \in \mathfrak{F}} C(y, f(X, \hat{\theta}_f)), \qquad (2)$$

причому оцінки параметрів для кожної моделі $f \in \mathfrak{F}$ є розв'язком ще однієї екстремальної задачі виду

$$\hat{\theta}_f = \arg \min_{\theta_f \in R^{s_f}} Q(y, X, \theta_f) \qquad (3)$$

де s_f називається складністю моделі f і дорівнює кількості ненульових компонентів у моделі виду (3); Q- критерій якості розв'язку задачі параметричної ідентифікації кожної окремої моделі, що генерується в задачі структурної ідентифікації.

МГУА володіє певним розмаїттям можливостей на всіх етапах процесу моделювання складних систем у порівнянні з іншими методами побудови моделей. Це стосується перш за все генераторів моделей і застосовуваних критеріїв якості структур, а також класів моделей (базисних функцій). Метод відрізняється активним застосуванням принципів автоматичної генерації

варіантів, послідовної селекції моделей і зовнішніх критеріїв для побудови моделей оптимальної складності. Він має оригінальну процедуру багаторядної автоматичної генерації структур моделей, що імітує процес біологічної селекції з попарним урахуванням послідовних ознак. Така процедура в сучасній термінології називається поліноміальною нейронною мережею, причому її структура є явною і будується автоматично, в режимі самоорганізації.

Для порівняння і вибору кращих моделей застосовуються зовнішні критерії, засновані на розділенні вибірки на дві й більш частин, причому оцінювання параметрів і перевірка якості моделей виконується на різних підвибірках. Це дозволяє обійтися без обтяжливих апріорних припущень, оскільки поділ вибірки дозволяє неявно (автоматично) врахувати різні види апріорної невизначеності при побудові моделі. МГУА має перевагу при малих вибірках за рахунок вибору складності моделі, що оптимально враховує інформативність наявних даних.

Ефективність методу багато разів підтверджувалася розв'язанням безлічі конкретних задач з областей екології, економіки, гідрометеорології тощо [47-49]. Теоретичні аспекти МГУА розглянуто в [50], [51]. Зокрема, в [50] на основі аналогії між задачею побудови моделі за зашумленими експериментальними даними і задачею проходження сигналу через канал з шумом побудовані начала теорії завадостійкого моделювання. Основний результат цієї теорії полягає в тому, що складність оптимальної прогнозуючої моделі залежить від рівня невизначеності в даних: чим він вище - тим простішою (більш грубою) має бути оптимальна модель (тим менше оцінюваних параметрів).

МГУА добре відомий і дуже активно розвивається у нас в країні й за кордоном [52-54]. Розроблено основи теорії структурної ідентифікації моделей з мінімальною дисперсією помилки прогнозування [50, 52-54]. Ефективним апаратом цієї теорії є метод критичних дисперсій, що вперше дозволяє аналітично розв'язувати актуальні задачі: порівняльний аналіз критеріїв структурної ідентифікації, планування експериментів, аналіз властивостей

методів тощо, причому як для обмеженої вибірки, так і в асимптотиці. При цьому досліджуються умови вибору оптимальної структури моделі залежно від дисперсії (рівня) шуму, довжини вибірки, вхідних впливів (плану експерименту) і параметрів об'єкта, причому встановлено тісний взаємозв'язок між ними. Засобами цієї теорії встановлено, що МГУА є методом побудови моделей з мінімальною дисперсією помилки прогнозування, і виконано порівняння його ефективності з іншими методами.

З цього випливає, що МГУА як основний інструмент теорії індуктивного моделювання належить до найсучасніших методів обчислювального інтелекту і м'яких обчислень. Цей метод є оригінальним і ефективним засобом розв'язання широкого спектру задач штучного інтелекту, в тому числі ідентифікації та прогнозування, розпізнавання образів і кластеризації, інтелектуального аналізу даних і пошуку закономірностей.

В останнє десятиліття інтерес до МГУА активно зростає в усьому світі, що можна пояснити, окрім відомої ефективності методу, також зростанням популярності технології штучних нейромереж. Річ у тім, що структуру МГУА можна інтерпретувати як нейромережу, оригінальність якої полягає в самоорганізації як її структури, так і параметрів. При цьому виявляється, що до явних переваг МГУА належать автоматичне формування структури мережі, простота і швидкодія настроювання параметрів, а також можливість «згорнути» побудовану мережу безпосередньо в явний математичний вираз.

Підтвердженням популярності МГУА є міжнародні форуми з індуктивного моделювання. Так, в період з 23 по 26 вересня 2007 р. в Празі було проведено II Міжнародний семінар з індуктивного моделювання (MCIM-2007) [57-58]. Перший семінар відбувся в Києві в липні 2005 р. як продовження Міжнародної конференції з індуктивного моделювання (МКІМ'2002) у Львові в травні 2002 р. Серія таких конференцій і семінарів - це міжнародні заходи, в центрі уваги яких стоять теорія, алгоритми, застосування, реалізації та нові розробки технологій аналізу даних і видобування знань, що базуються на методології МГУА.

Підхід індуктивного моделювання, побудований на принципах самоорганізації, активно розвивається протягом 40 років, застосовується в багатьох областях і присутній в таких поширених технологіях аналізу даних, як поліноміальні нейронні мережі, адаптивні та статистичні мережі, що навчаються. В нових розробках для побудови моделей на основі даних використовуються також еволюційні й генетичні алгоритми, ідея активних нейронів і багаторівнева самоорганізація, та інші ідеї.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Darwin, C. (1859). On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life. London: J. Murray;
2. Galton, Francis (1886). "Regression Towards Mediocrity in Hereditary Stature". *The Journal of the Anthropological Institute of Great Britain and Ireland*. 15: 246–263;
3. Stigler, Stephen M. (1 July 2010). "Darwin, Galton and the Statistical Enlightenment". *Journal of the Royal Statistical Society, Series A*. 173;
4. Francis Galton (1886) Anthropological Miscellanea: "Regression towards mediocrity in hereditary stature," *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15: 246–263;
5. Linear Regression – Wikipedia [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Linear_regression
6. Ohm's law – Wikipedia [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Ohm%27s_law
7. Least squares – Wikipedia [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Least_squares
8. Лежандр, Адриен Мари (1805). “Нові методи для визначення орбіт комет” London: J. Murray;
9. Р. Адрейн (1808), “Дослідження, що стосуються ймовірностей помилок, які зє при змінах”, Accademia delle Scienze di Torino. pp. 86–95

10. Stephen M. Stigler, "Gauss and the Invention of Least Squares," *Ann. Statist.*, 9 (3), 1981, pp. 465–474.
11. "Pierre Simon, Marquis De Laplace" (1911). Encyclopædia Britannica
12. Bruno, Leonard C. (2003) [1999]. Math and mathematicians : the history of math discoveries around the world. Baker, Lawrence W. Detroit, Mich.: U X L. p. 67. ISBN 0787638137. OCLC 41497065.
13. A.A. Markov. "Extension of the limit theorems of probability theory to a sum of variables connected in a chain". reprinted in Appendix B of: R. Howard. Dynamic Probabilistic Systems, volume 1: Markov Chains. John Wiley and Sons, 1971.
14. James H. Stock, Mark W. Watson. Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals // Introduction to Econometrics. — 3. — Addison-Wesley, 2011. — P. 163-164. — 785 p. — ISBN 0138009007.
15. Merriman M., Transaction of the Connecticute Academy of Arts and Science, 4, 1877.
16. Д. Худсон. Статистика для физиков. Москва, Мир, 1970
17. Norman Richard Draper, Harry Smith. “Applied Regression Analysis” Wiley, 1998.
18. *Chihara, Theodore Seio (1978). An Introduction to Orthogonal Polynomials. Gordon and Breach, New York.*
19. *Jackson, Dunham (2004) [1941]. Fourier Series and Orthogonal Polynomials. New York: Dover.*
20. D. Van Der Reyden, “Curve Fitting by the Orthogonal Polynoms of Least Squares”, 1943.
21. Gram–Schmidt process – Wikipedia [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Gram%E2%80%93Schmidt_process
22. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений / Пер. с англ. Х.Д. Икрамова. – М.: Мир, 1980. – 280 с.

23. Forsythe G, journ. Sos.Ind. Appl. Math 5,74 (1957)
24. Weatherburn C., A first course in mathematical statistics, Combridge, 1952.
25. Devies O., Statistical methods in research adn production, New York, 1957.
26. Plackett R., Principles of regression analysis, Oxford, 1960.
27. Kendall M. G., Stuart A., The advanced theory of statistics, vil. 1, 2, New York, 1958.
28. Williams E., Regression Analysis, 1959.
29. Jauneau J., Morellet D., Procedeeing of the 1964 Easter School for Physicists, vol. 1, 1 1964, CERN.
30. Grendander U., Rosenblatt M., Statistical analysis of stationary time series, New York, 1957.
31. L1 and L2 Regulatization methods [Електронний ресурс] – Режим доступу до ресурсу: https://towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c
32. Regularization (Mathematics) - Wikipedia [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Regularization_(mathematics)
33. Avoiding overfitting with regulatization [Електронний ресурс] – Режим доступу до ресурсу: https://www.analyticsvidhya.com/blog/2015/02/avoid-over-fitting-regularization/
34. A. Neumaier, Solving ill-conditioned and singular linear systems: A tutorial on regularization, SIAM Review 40 (1998), 636–666.
35. Local Regression – Wikipedia [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Local_regression
36. Cleveland, William S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots". *Journal of the American Statistical Association*. 74 (368): 829–836
37. https://en.wikipedia.org/wiki/Multivariate_adaptive_regression_splines

38. Friedman, J. H. (1991). "Multivariate Adaptive Regression Splines". The Annals of Statistics. 19: 1.
39. МЗ Згуровский, АА Павлов, "Принятие решений в сетевых системах с ограниченными ресурсами: Монография", К.: Наукова думка,–2010.–573 с
40. Згуровский М.З., Павлов А.А., Мисюра Е.Б., Мельников О.В. Методы оперативного планирования и принятия решений в сложных организационно-технологических системах // Вісник НТУУ “КПІ”. Інформатика, управління та обчислювальна техніка. К.: “ВЕК+”, 2010.– No50
41. D. J. FYFE, K. H. OKE; Use of Residuals in Forsythe's Method for Polynomial Curve Fitting, *IMA Journal of Applied Mathematics*, Volume 24, Issue 1, 1 August 1979, Pages 99–108
42. Павлов А.А., Калашник В.В., Коваленко Д.А. Построение багатовимірної поліномиальной регресії. Регресія при даних з повторюваними аргументуми // Вісник НТУУ “КПІ”. Серія «Інформатика, управління та обчислювальна техніка». – К.: “ВЕК+”, 2015. – No63. – 4 с.
43. Згуровский М.З., Павлов А.А. Принятие решений в сетевых системах с ограни- ченными ресурсами: Монография. – К.: Наукова думка. – 2010. – 573 с.
44. “Learning representations by back-propagating errors.” Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams.
45. Aleksandar Peckov, "A MACHINE LEARNING APPROACH TO POLYNOMIAL REGRESSION", Doctoral Dissertation, Jozef Stefan International Postgraduate School, 2012.
46. Ивахненко А.Г. Метод групового урахування аргументів - конкурент методу стохастичної апроксимації // Автоматика. - 1968. - № 3. - С. 58-72.
47. Ивахненко А.Г. Системи эвристической самоорганизации в технической кибернетике. - Киев: "Техніка", 1971. - 392 с.
48. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами. - Киев: "Техніка", 1975. - 311 с.

49. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем. - Киев: "Наук. думка", 1982. - 296 с.
50. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделірования. - Киев: "Наук. думка", 1985. - 216 с.
51. Ивахненко А.Г., Юрачковский Ю.П. Моделірование сложных систем по експериментульным данным. - М.: "Радио и звязок", 1987. - 120 с.
52. Ivachnenko A.G., Muller J.A. Selbstorganisation von Vorhersagemodellen. - Berlin: Veb Verlag Technik, 1984. - 223 c.
53. Self-organizing methods in modeling: GMDH type algorithms / Ed. S.J.Farlow. - New York, Basel: Marcel Decker Inc., 1984. - 350 p.
54. Madala, H. R., Ivakhnenko, A.G. Inductive learning algorithms for complex systems modeling. - New York: Boca Raton, CRC Press, 1994. - 384 c.
55. Степашко В.С. Аналіз ефективности критериев структурной идентификации прогнозирующих моделей // Проблемы управления и информатики. - 1994. - № 3-4. - С. 13-21.
56. Степашко В.С. Структурна ідентифікація моделей як задача відновлення сигналу в умовах неповноти інформації // Наукові праці ДНТУ. Серія: Обчисл. техніка та автоматизація. - Вип. 48. - Донецьк: ДНТУ, 2002. - С. 345-353.
57. Степашко В.С. Теоретичні аспекти МГУА як методу індуктивного моделювання // Управляющие системи и машины. - 2003. - №2. - с.31-44.
58. Proceedings of International Workshop on Inductive Modelling (IWIM 2007). - Prague: Czech Technical University, 2007. - 329 p. - ISBN 978-80-01-03881-

2 ПОСТАНОВКА ТА ФОРМАЛІЗАЦІЯ ЗАДАЧІ

2.1 Змістовна постановка задачі

Проблема знаходження істинної закономірності по результатам вимірів (вхід-вихід) є універсальною. Немає жодної області діяльності людини, в якій так чи інакше ця задача не виникала. Різноманітні аспекти цієї проблеми розглядаються в таких науках, як математична статистика, теорія управління, теорія штучного інтелекту та інші. В рамках ймовірнісних моделей ця задача формулюється як оцінка лінії регресії по результатам статистичних експериментів і в практичному плані являє собою область застосування прикладного регресійного аналізу.

Проблема відтворення невідомої закономірності формулюється як класична задача прикладного регресійного аналізу: відтворення багатовимірної поліноміальної регресії по надлишковому опису і довільно розподіленою похибкою. В такій постановці задача виникає, коли вигляд закономірності точно не відомий, але, виходячи зі знань про об’єкт, дослідник здатен задати її опис, який, можливо, містить надлишкові члени. Сам надлишковий опис задається скінченним багатовимірним поліномом від незалежних змінних з невідомими коефіцієнтами, підмножина яких є нулями. По результатам статистичних експериментів необхідно оцінити невідомі коефіцієнти, частина із яких тотожно дорівнює нулю і невідома досліднику. Пропонується і обґрунтовується ефективний алгоритм вирішення даної задачі, що базується на використанні ортогональних поліномів Форсайта в рамках класичного методу найменших квадратів. Основна прикладна особливість запропонованого методу – відсутність обчислювальної складності, формування вимог до експериментів, виконання яких приводить до практично точному відновленню закономірності багатовимірної поліноміальної регресії при адитивній похибці з скінченною дисперсією.

2.2 Математична постановка задачі

Завдання конструктивного відновлення по статистичним даним регресійної моделі (детермінованої закономірності) є предметом вивчення прикладного регресійного аналізу [1, 2, 11, 27, 36, 68, 80]. Найбільш широко використовуваним методом є метод найменших квадратів. Практичні проблеми реалізації методу найменших квадратів, при створенні багатовимірного полінома регресії є необхідність обернення погано детермінованих матриць; відсутність ефективної процедури для відновлення істинної багатовимірної поліноміальної регресії по надлишковому опису. Запропонований метод в цілому ефективно справляється з обома проблемами. Основою його побудови сформульовані в [57, 58, 59].

2.2.1 Модель одновимірної поліноміальної регресії

Постановка задачі та аналіз з відомими результатами відповідно до [9]. Модель регресії має вигляд

$$Y(x) = \theta_0 + \theta_1 x + \ldots + \theta_r x^r + \mathrm{E}, \tag{0.1}$$

де x – детермінована змінна, значення якої в експериментах дослідник може задавати довільно; θ_i, $i = \overline{0,r}$ – невідомі коефіцієнти, E – випадкова величина з довільним розподілом, $ME = 0$ (M – знак математичного очікування); σ_E^2 (дисперсія) обмежена, її значення невідомо, або відома її верхня оцінка σ^2.

Проведено n експериментів, результатом яких є дві вибірка розміру n $\left(x_i,\ i = \overline{1,n};\ Y(x_i) = y_i,\ \ i = \overline{1,n}\right)$.

Відповідно до (0.1)

$$y_i = \sum_{j=0}^{r} \theta_j x_i^j + \delta_i, \quad i = \overline{1,n}; \tag{0.2}$$

де δ_i – невідома реалізація випадкової величини E в i-тому експерименті. Числа y_i, δ_i можна вважати реалізаціями випадкових величин Y_i $i = \overline{1,n}$; Δ_i $i = \overline{1,n}$, де Δ_i має розподіл випадкової величини E, а Y_i i Δ_i зв'язаних співвідношенням:

$$Y_i = \sum_{j=0}^{r} \theta_j x_i^j + \Delta_i, \tag{0.3}$$

де Δ_i, $i = \overline{1,n}$ – незалежні випадкові величини, розподілені так само, як і випадкова величина E; Y_i $i = \overline{1,n}$ – незалежні випадкові величини з дисперією σ_E^2

Оцінки невідомих коефіцієнтів θ_j, $j = \overline{0,r}$ знаходяться із мінімізації виразу

$$\min_{\theta_j, j=0,r} \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{r} \theta_j x_i^j \right)^2 \tag{0.4}$$

Введемоо матричні позначення:

$$A = \begin{pmatrix} 1x_1 \ldots x_1^r \\ \cdots\cdots\cdots \\ 1x_n, \ldots x_n^r \end{pmatrix}; \quad y = \left(y_1, \ldots y_n\right)^{\mathsf{T}}$$

$$Y = \left(Y_1, \ldots Y_n\right)^{\mathsf{T}}; \quad \theta = \left(\theta_0, \ldots \theta_r\right)^{\mathsf{T}}$$

$$\hat{\theta} = \left(\hat{\theta}_1, \ldots \hat{\theta}_r\right)^{\mathsf{T}},$$

де $\hat{\theta}_j$ – оцінки θ_j, $j = \overline{0,r}$ відповідно до (0.4).

Тоді [**Error! Reference source not found.**]

$$\hat{\theta} = \left(A^{\mathsf{T}} A\right)^{-1} A^{\mathsf{T}} y, \tag{0.5}$$

або $\hat{\theta} = \left(A^{\mathsf{T}} A\right)^{-1} A^{\mathsf{T}} Y$, якщо $\hat{\theta}_j$, $j = \overline{0,r}$, вважати випадковими величинами.

2.2.2 Модель багатовимірної поліноміальної регресії

Нехай багатовимірна регресія задається у вигляді

$$y\left(\bar{x}\right) = \sum_{\forall (i_1, \ldots i_t) \in \mathrm{K}} \sum_{\forall (j_1, \ldots j_t) \in \mathrm{K}(i_1, \ldots i_t)} b_{i_1 \ldots i_t}^{j_1 \ldots j_t} \left(x_{i_1}\right)^{j_1} \cdot \left(x_{i_2}\right)^{j_2} \ldots \left(x_{i_t}\right)^{j_t} + E, \tag{0.6}$$

де $\bar{x} = \left(x_1 \ldots x_n\right)^{\mathsf{T}}$ – детермінований вектор вхідних змінних, $\tilde{o}_i$ – i-та компонента вектора $\bar{x}$; $b_{i_1 \ldots i_t}^{j_1 \ldots j_t}$ – невідомі коефіцієнти, j_l – натуральні числа; i_l – натуральні індекси із множини $\{1, \ldots, n\}$; $\hat{A}$ – випадкова величина з нульовим математичним очікуванням и обмеженою невідомою дисперсією σ_E^2 (як и в одновимірному випадку може бути відома верхняоцінка σ_E^2).

Модель (0.6) є надлишковою – можливо, деякі із коефіцієнтів $b_{i_1\ldots i_t}^{j_1\ldots j_t}$ рівні нулю. Для зручності подальшого викладу лінію регресії моделі (0.6) представимо інакше:

$$\sum_{l=1}^{n} \sum_{\forall(i_1,\ldots i_t)\in \mathrm{K}_l} \sum_{\forall(j_1,\ldots j_t)\in \mathrm{K}_l(i_1,\ldots i_t)} b_{i_1\ldots i_t}^{j_1\ldots j_t} \left(x_{i_1}\right)^{j_1} \cdot \left(x_{i_2}\right)^{j_2} \ldots \left(x_{i_t}\right)^{j_t} \tag{0.7}$$

Складові

$$\sum_{\forall(i_1,\ldots i_t)\in \mathrm{K}_1} \sum_{\forall(j_1,\ldots j_t)\in \mathrm{K}_1(i_1,\ldots i_t)} b_{i_1\ldots i_t}^{j_1\ldots j_t} \left(x_{i_1}\right)^{j_1} \cdot \left(x_{i_2}\right)^{j_2} \ldots \left(x_{i_t}\right)^{j_t} \tag{0.8}$$

містять всі елементи із (0.6), в кожну із яких входить компонента x_1. Складові

$$\sum_{\forall(i_1,\ldots i_t)\in \mathrm{K}_l} \sum_{\forall(j_1,\ldots j_t)\in \mathrm{K}_l(i_1,\ldots i_t)} b_{i_1\ldots i_t}^{j_1\ldots j_t} \left(x_{i_1}\right)^{j_1} \cdot \left(x_{i_2}\right)^{j_2} \ldots \left(x_{i_t}\right)^{j_t} \quad l = \overline{2,n} \tag{0.9}$$

містять всі елементи із (0.6), в кожну із яких входить компонента x_l, за вийнятком тих елементів, які увійшли до (0.8) і (0.9) для

$$\forall\left(i_1,\ldots i_t\right)\in \mathrm{K}_m \forall\left(j_1,\ldots j_t\right)\in \mathrm{K}_m\left(i_1,\ldots t_t\right), m = \overline{1,l-1}.$$

2.3 Обґрунтування методу розв'язання

Пропонується і обґрунтовується ефективний алгоритм вирішення задачі багатовимірного регресійного аналізу, що базується на використанні нормованих поліномів Форсайта [10] в рамках класичного методу найменших квадратів. Основна прикладна особливість методу, запропонованого нижче – відсутність обчислювальної складності, формування вимог до експерименту (плану експерименту), виконання яких призводить до майже точного відтворення закономірності (багатовимірної поліноміальної регресії) при адитивній похибці з довільною скінченною похибкою.

В [11] приведений метод побудови багатовимірної поліноміальної регресії по надлишковому опису в умовах активного експерименту. Алгоритм базується на зведенні багатовимірної регресії до одновимірної за допомогою фіксації всіх змінних окрім однієї. Цей метод має той недолік, що в одновимірних регресіях, що будуються в процесі роботи алгоритму, міститься велика кількість членів з

малими значеннями ступенів при *x*. Як буде показано нижче, коефіцієнти таких поліномів погано відтворюються.

В даній роботі пропонується модифікований алгоритм побудови багатовимірної поліноміальної регресії, що базується на ідеях, викладених в [12]. Основна ідея модифікованого методу полягає в фіксації невеликої кількості змінних та одночасній зміні інших (не фіксованих) параметрів. Це призводить до одновимірних регресій з вищими степенями та більш точними оцінками невідомих коефіцієнтів.

2.4 Опис методів розв'язання

2.4.1 Знаходження одновимірної лінії регресії з рекурентними співвідношеннями Форсайта

Складності, зв'язані з оберненням матриці $\left(A^{\top}A\right)^{-1}$ зникають, якщо від моделі (0.1) перейти до моделі регресії, заданої за допомогою ортогональних поліномів[**Error! Reference source not found.**]:

$$Y(x) = w_0\theta_0(x) + w_1\theta_1(x) + \ldots + w_r\theta_r(x) + E, \tag{0.10}$$

де $\theta_j(x)$, $j = \overline{0,r}$ – нормовані ортогональні поліноми,

$$\theta_j(x) = q_{j_0} + q_{j_1}x + \ldots + q_{jj}x^j \tag{0.11}$$

$$\sum_{i=l}^{n}\theta_j^2(x_i) = 1,\ \sum_{i=l}^{n}\theta_j(x_i)\theta_l(x_i) = 0\ \forall j \neq l,\ j,l = \overline{0,r}$$

Дж. Форсайт [**Error! Reference source not found.**] запропонував рекурентну формулу для ззнаходження нормованих ортогональних поліномів:

$$\lambda\,\theta_j(x) = x\,\theta_{j-1}(x) - \alpha\,\theta_{j-1}(x) - \beta\,\theta_{j-2}(x) \tag{0.12}$$

$$\alpha = \sum_{i=1}^{n} x_i\,\theta_{j-1}^2(x_i);\ \beta = \sum_{i=1}^{n} x_i\,\theta_{j-1}(x_i)\theta_{j-2}(x_i);$$

λ визначається із умови $\sum_{i=1}^{n}\theta_j^2(x_i) = 1.$

$$\lambda = \sqrt{\sum_{i-1}^{n}\left(x_i\theta_{j-1}(x_i) - \alpha\theta_{j-1}(x_i) - \beta\theta_{j-2}(x_i)\right)^2}.$$

Для використання рекурентної формули (0.12) необхідно побудувати нормовані ортогональні поліноми $\theta_0(x)$ *и* $\theta_1(x)$. Очевидно, ними є

$$\theta_0(x)=\frac{1}{\sqrt{n}};\quad \theta_1(x)=-\frac{\bar{x}}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}}+\frac{x}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}},$$

де $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$.

Застосування методу найменших квадратів до моделі (0.10) призводить до наступних результатів[**Error! Reference source not found.**]:

Нехай $w=(w_0,...w_r)^{\mathsf{T}}$; $\hat{w}=(\hat{w}_0,...\hat{w}_r)^{\mathsf{T}}$ $\hat{w}_j, j=\overline{0,r}$ оцінки w_j отримані методом найменших квадратів $\hat{W}=\left(\hat{W}_0,...\hat{W}_r\right)^{\mathsf{T}}$, $\hat{W}_j, j=\overline{0,r}$ – випадкові величини, для яких $\hat{w}_j$ є відповідними реалізаціями. Тоді

$$\hat{w}_j=\sum_{i=1}^{n}y_i\theta_j(x_i),\quad j=\overline{0,r};\quad \hat{W}_j=\sum_{i=1}^{n}Y_i\theta_j(x_i) \tag{0.13}$$

$$M\hat{W}_j=w_j,\ j=\overline{0,r};\ \operatorname{cov}\left(\hat{W}_j,\ \hat{W}_l\right)=0,\ \forall\, j\neq l.$$

$$D\hat{W}_j=\sigma_E^2 \tag{0.14}$$

$$M\frac{\sum_{i=1}^{n}Y_i^2-\sum_{j=0}^{r}\hat{W}_j^2}{n-(r+1)}=\sigma_E^2 \tag{0.15}$$

Покажемо, наприклад, що $\operatorname{cov}\left(\hat{W}_j,\ \hat{W}_l\right)=0,\ \forall\, j\neq l$ (це доведення відсутнє в [**Error! Reference source not found.**]). $\operatorname{cov}\left(\hat{W}_l,\ \hat{W}_p\right)=0,\ l\neq p,$ якщо $M\left(\hat{W}_l\cdot\hat{W}_p\right)=M\hat{W}_l\cdot M\hat{W}_p$:

$$\hat{W}_l=\sum_{i=1}^{n}Y_i\theta_l(x_i),\quad \hat{W}_p=\sum_{i=1}^{n}Y_i\theta_p(x_i),\ M\hat{W}_l=\sum_{i=1}^{n}\theta_l(x_i)\cdot MY_i,\quad M\hat{W}_p=\sum_{j=1}^{n}\theta_p(x_j)\cdot MY_j,$$

$$M\left(\hat{W}_l\cdot\hat{W}_p\right)=\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\theta_l(x_i)\theta_p(x_j)\cdot MY_i\cdot MY_j+\sum_{i=1}^{n}M\left(Y_i\right)^2\cdot\theta_l(x_i)\theta_p(x_i) \tag{0.16}$$

$$M\hat{W}_l\cdot M\hat{W}_p=\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\theta_l(x_i)\theta_p(x_j)\cdot MY_i\cdot MY_j+\sum_{i=1}^{n}\left(MY_i\right)^2\cdot\theta_l(x_i)\theta_p(x_i). \tag{0.17}$$

Розглянемо різницю (0.16) і (0.17):

$$M\left(\hat{W}_l \cdot \hat{W}_p\right) - M\hat{W}_l \cdot M\hat{W}_p = \sum_{i=1}^{n}\left(M\left(Y_i\right)^2 - \left(M\,Y_i\right)^2\right)\cdot\theta_l\left(x_i\right)\theta_p\left(x_i\right) =$$

$$= \sum_{i=1}^{n} DY_i \cdot \theta_l\left(x_i\right)\theta_p\left(x_i\right) = \sigma^2 \sum_{i=1}^{n} \theta_l\left(x_i\right)\theta_p\left(x_i\right) = 0,$$

при $l \neq p$, так як $\theta_l\left(x\right)$ і $\theta_p\left(x\right)$ – ортогональні поліноми.

Звязок моделей (0.1) і (0.10) є наступний:

$$\theta_j = w_r q_{rj} + w_{r-1} q_{r-1,j} + \ldots + w_j q_{jj} \tag{0.18}$$

і, відповідно,

$$\hat{\theta}_j = \hat{w}_r q_{rj} + \ldots + \hat{w}_j q_{jj}, j = \overline{0,r} \tag{0.19}$$

або

$$\hat{\theta}_j = \hat{W}_r q_{rj} + \ldots + \hat{W}_j q_{jj}, \tag{0.20}$$

якщо $\hat{\theta}_j$ вважати випадковою величиною.

При дослідженні моделі (0.1) або їй еквівалентній (0.10) в [**Error! Reference source not found.**] передбачалося, що r – степінь полінома регресії відома заздалегіть. Якщо це не так, то прийнято вважати [**Error! Reference source not found.**], що для випадкового розподілу E знаходження істинного r є проблемою. Якщо E має нормальний розподіл, то знаходження r зводиться до перевірки стетичтичних гіпотез по критеріям з відомим розподілом Фішера [**Error! Reference source not found.**].

Покажемо, що насправді проблема ззнаходження r має конструктивне вирішення для довільного розподілу E. Також покажемо, як можна ефективно звязати наявні експериментальні дані з точністю оцінок невідомих коефіцієнтів $\theta_j, \quad j = \overline{0,r}$.

Умова $\sum_{i=1}^{n} \theta_j^2\left(x_i\right) = 1$ з врахуванням (0.11) перепишемо наступним чином:

$$\sum_{i=1}^{n}\left(\sum_{l=0}^{j} q_{jl} x_i^l\right)^2 = 1 \tag{0.21}$$

Знайдем дисперсію $\hat{\theta}_j$. Враховуючи, що $\operatorname{cov}\left(\hat{W}_l, \hat{W}_p\right) = 0$, $l \neq p$ із (0.14) і (0.20) отримаємо

$$D\hat{\theta}_j = \sigma^2 \sum_{l=r}^{j} q_{lj}^2 \tag{0.22}$$

Так як при необмеженому зростанні числа випробувань n мінімум (0.4) асимптотично повинен досягатися на істинних значеннях коефіцієнтів θ_j, із аналізу (0.21) і (0.22) слідує, що при збільшенні n модулі значень коефіцієнтів $\left|q_{lj}\right|$, $l = \overline{r, j}$, $j = \overline{0, r}$ повинні зменшитись.

Аналогічно в загальному вигляді досить складно звязати числа x_i, $i = \overline{1, n}$; n; $j\left(j = \overline{0, r}\right)$ з величиною q_{lj}, $l = \overline{r, j}$, $j = \overline{0, r}$. Тим не менш, у випадку активного експерименту для ефективного вирішення прикладних задач (задана точність; необхідна кількість обчислень; визначення чисел $x_1, \ldots, x_n$) можна створити відповідні статистичні таблиці, можливий фрагмент однієї із них представлений у табл. 0.1.

Таблиця побудована для ліній регресії заданої поліномом пятого порядку. В першому стовбці фіксуються різні значення n (кількість значень детермінованого аргументу x). В стовбцях з номером $j\left(j = \overline{0,5}\right)$ задані дисперсії коефіцієнтів $\hat{\theta}_j$, $j = \overline{0,5}$, як функція σ^2 (σ^2 – це дисперсія E або її верхня оцінка). Для побудови таблиці були знайдені всі нормовані ортогональні поліноми $\theta_j(x)$, $j = \overline{0,5}$ (використовуются формули (0.11), (0.12)), по (0.22) визначені відповідні дисперсії. Значення x_i, $i = \overline{1, n}$, розподілені з однаковим кроком по відрізку (–50,0, 50,0)

Таблиця 0.1 – Фрагмент можливої ситуації

n	0	1	2	3	4	5
10	$\sigma^2 \cdot 0{,}400466$	$\sigma^2 \cdot 0{,}0024855$	$\sigma^2 \cdot 4{,}26 \cdot 10^{-06}$	$\sigma^2 \cdot 7{,}55 \cdot 10^{-09}$	$\sigma^2 \cdot 1{,}41 \cdot 10^{-12}$	$\sigma^2 \cdot 1{,}28 \cdot 10^{-15}$
50	$\sigma^2 \cdot 0{,}0706426$	$\sigma^2 \cdot 0{,}0004607$	$\sigma^2 \cdot 4{,}53 \cdot 10^{-07}$	$\sigma^2 \cdot 1{,}15 \cdot 10^{-09}$	$\sigma^2 \cdot 9{,}28 \cdot 10^{-14}$	$\sigma^2 \cdot 1{,}43 \cdot 10^{-16}$
100	$\sigma^2 \cdot 0{,}0351973$	$\sigma^2 \cdot 0{,}0002298$	$\sigma^2 \cdot 2{,}22 \cdot 10^{-07}$	$\sigma^2 \cdot 5{,}68 \cdot 10^{-10}$	$\sigma^2 \cdot 4{,}47 \cdot 10^{-14}$	$\sigma^2 \cdot 7{,}02 \cdot 10^{-17}$
200	$\sigma^2 \cdot 0{,}0175833$	$\sigma^2 \cdot 0{,}0001149$	$\sigma^2 \cdot 1{,}10 \cdot 10^{-07}$	$\sigma^2 \cdot 2{,}84 \cdot 10^{-10}$	$\sigma^2 \cdot 2{,}21 \cdot 10^{-14}$	$\sigma^2 \cdot 3{,}50 \cdot 10^{-17}$
300	$\sigma^2 \cdot 0{,}0117203$	$\sigma^2 \cdot 7{,}66 \cdot 10^{-05}$	$\sigma^2 \cdot 7{,}36 \cdot 10^{-08}$	$\sigma^2 \cdot 1{,}89 \cdot 10^{-10}$	$\sigma^2 \cdot 1{,}47 \cdot 10^{-14}$	$\sigma^2 \cdot 2{,}33 \cdot 10^{-17}$
500	$\sigma^2 \cdot 0{,}0070316$	$\sigma^2 \cdot 4{,}59 \cdot 10^{-05}$	$\sigma^2 \cdot 4{,}41 \cdot 10^{-08}$	$\sigma^2 \cdot 1{,}13 \cdot 10^{-10}$	$\sigma^2 \cdot 8{,}82 \cdot 10^{-15}$	$\sigma^2 \cdot 1{,}40 \cdot 10^{-17}$
1000	$\sigma^2 \cdot 0{,}0035157$	$\sigma^2 \cdot 2{,}30 \cdot 10^{-05}$	$\sigma^2 \cdot 2{,}21 \cdot 10^{-08}$	$\sigma^2 \cdot 5{,}67 \cdot 10^{-11}$	$\sigma^2 \cdot 4{,}41 \cdot 10^{-15}$	$\sigma^2 \cdot 6{,}99 \cdot 10^{-18}$
5000	$\sigma^2 \cdot 0{,}0007031$	$\sigma^2 \cdot 4{,}59 \cdot 10^{-06}$	$\sigma^2 \cdot 4{,}41 \cdot 10^{-09}$	$\sigma^2 \cdot 1{,}13 \cdot 10^{-11}$	$\sigma^2 \cdot 8{,}82 \cdot 10^{-16}$	$\sigma^2 \cdot 1{,}40 \cdot 10^{-18}$
10000	$\sigma^2 \cdot 0{,}0003516$	$\sigma^2 \cdot 2{,}30 \cdot 10^{-06}$	$\sigma^2 \cdot 2{,}21 \cdot 10^{-09}$	$\sigma^2 \cdot 5{,}67 \cdot 10^{-12}$	$\sigma^2 \cdot 4{,}41 \cdot 10^{-16}$	$\sigma^2 \cdot 6{,}99 \cdot 10^{-19}$

На якісному рівні аналіз табл. 0.1 не залежить від величини $a > 1$ відрізка розбиття ($-a$, a) і величин r – степені полінома. Викладені нижче висновки підтверджені експериментально.

1. Приведені значення дисперсій $\hat{\theta}_j$, $j = \overline{0,5}$, є конструктивними, якщо відома верхня оцінка σ^2 дисперсії E. Порядок σ_E^2 можна визначити по реалізації випадкової величини [**Error! Reference source not found.**]:

$$\frac{R^T R}{n-(r+1)} = \frac{\sum_{i-1}^{n} y_i^2 - \sum_{j-0}^{r} W_j^2}{n-(r+1)}, \text{ так як } M\frac{R^T R}{n-(r+1)} = \sigma_E^2$$

Далі буде показано, що істинне значення r знаходять очевидним чином.

2. Чим більше j, тим менше $D\hat{\theta}_j$ при фіксованому n. Дійсно, при $n = 10$ $D\hat{\theta}_0 = \sigma^2 \cdot 0{,}400466,\ D\hat{\theta}_1 = \sigma^2 \cdot 0{,}0024855,\ D\hat{\theta}_2 = \sigma^2 \cdot 4{,}26 \cdot 10^{-6} \ldots D\hat{\theta}_5 = \sigma^2 \cdot 128 \cdot 10^{-15}$ тобто с увеличением j значення $D\hat{\theta}_j$ зменшується на порядок. Із цього слідує наступний результат.

3. По мінімальній кількості випробувань можна визначити істинну ступінь полінома лінії регресії в випадку, коли невідомі ненульові коефіцієнти θ_j не є малими по абсолютній величині числами. В нашому прикладі при $n = 10$ дисперсія оцінки коефіцієнта при x^2 вже рівна $\sigma^2 \cdot 4{,}26 \cdot 10^{-6}$, тобто якщо істинна лінія регресії пряма, то насправді оцінками $\hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_5$ будуть нулі з точністю до відповідних знаків після коми (закон трьох сігм для нормального розводілу і використання нерівності Чебишева в загальному випадку).

4. Необхідна кількість випробувань n визначається заданої точністю для знаходження $\hat{\theta}_j$ с найменшим $j\left(j=0\right)$. Якщо експерименти є дорогими, то реально ефективно оцінити $\hat{\theta}_j$ потрібно, починаючиз $j=1$ (із аналізу табл. 0.1 видно, що значення дисперсій $D\hat{\theta}_0$і $D\hat{\theta}_1$ одного порядка достигаются на числе експериментів отличающихся на два порядка).

Таким чином, точність оцінки θ_0 необхідно связыватьзотриманої числовой оценкой θ_0 (чем більше по модулю це значение, тем достовернее полученный результат). Якщо оцінка θ_0 оказывается недостатньо точной, то полученное выражение для линії регресії необхідно використовуватив тих задачах, для рішення яких величина θ_0 не має значення (Наприклад, сравнение значень линії регресії для різних значень її аргументу).

В деяких задачах масив $x_i,\ i=\overline{1,n}$, може бути заданий заздалегідь і експериментатор не може його змінити. Тоді до проведення експерименту по формулам (0.11), (0.12), (0.22) можна знайти дисперсії $\hat{\theta}_j,\ j=0,r$ (r можна задати надлишковим) і провести попередній аналіз майбутніх результатів експерименту.

2.2.2 Знаходження багатовимірної поліноміальної лінії регресії

Розглянемо складову (0.8). Позначемо через $\mathrm{M}_j^1,\ j=\overline{1,n_1}$, кількість доданків, кожен із яких містить x_1 в j-й ступені; $\mathrm{M}^1=\max\limits_j \mathrm{M}_j^1\ \ j=\overline{1,n_1}$, n_1 – максимальна ступінь полінома від змінної x_1.

Фіксуємо M^1 наборів значень компонент $\delta_2^s,\ldots,\delta_n^s\ \ s=\overline{1,M^1}$. На числа $x_i^s\ i=\overline{2,n},\ s=\overline{1,\mathrm{M}^1}$, накладається єдина умова – визначені нижче квадратні матриці повинні бути невиродженими.

Реалізуємо M^1 експериментів, в кожному із яких (s-м, $s=\overline{1,M^1}$) змінні $x_{2,}\ldots x_n$ приймають фіксовані значення $x_i^s\left(i=\overline{2,n}\right),$ а змінна x_1 змінюється як при побудові одновимірної поліноміальної регресії. При фіксованих значеннях

змінних $x_{2,}\ldots x_n$ в s-м експерименті $\left(s=\overline{1,M^1}\right)$ багатовимірна лінія регресії перетворюється в поліном від змінної x_1 ступені n_1.

Для кожного s-го експерименту $\left(s=\overline{1,M^1}\right)$ знаходимо значення дисперсій $D\widehat{\theta}_j^s$ (по виразу (0.22)), $j=\overline{1,n_1}$, і ці числа ранжуємо по зростанню їх значень при фиксированном j. Отримаємо n_1 проранжированных послідовательностей оцінок коефіцієнтів $\theta_j^{s_1},\ldots\theta_j^{s_{M^1}}\left(j=\overline{1,n_1}\right)$.

Ці результати дозволяють сформировать n_1 систем лінійних рівнянь, рішеннями яких є значення всіх коефіцієнтів $b_{i_1\ldots i_t}^{j_1\cdots j_t}$ в виразі (0.8).

Дійсно, в кожному із s експериментів невідомі коефіцієнти $\widehat{\theta}_j^s\left(j=\overline{1,n_1}\right)$ одновимірної поліномиальної регресії ступені n_1 від змінної x_1 определяются наступним чином: необхідно із всіх членів виразу (содержащих змінну x_1 в ступені j) вынести x_1^j. Полученное выражение для θ_j^s містить тільки M_j^1 невідомих коефіцієнтів виду $b_{i_1\ldots i_t}^{j_1\cdots j_t}$, так як в кожному s-м експерименті при зміні значень змінної x_1 змінні x_i, $i=\overline{2,n}$ приймають одноіто же фиксированное значення x_i^s, $(i=\overline{2,n})$. Таким чином, для побудови системи лінійних рівнянь для знаходження M_1^1 коефіцієнтіввиду $b_{i_1\ldots i_t}^{j_1\cdots j_t}$ надо використовувати M_1^1 чисел $\widehat{\theta}_1^{s_1},\ldots\widehat{\theta}_1^{s_{M_1^1}}$ (они мають наименьшую дисперсію).

Для визначення верхніх статистичних оцінок точності знаходження M_1^1 коефіцієнта виду $b_{i_1\ldots i_t}^{j_1\cdots j_t}$, отриману систему лінійних рівнянь запишемо так:

$$A\begin{pmatrix}x_1\\ \vdots\\ x_{M_1^1}\end{pmatrix}=\begin{pmatrix}\widehat{\theta}_1^{s_1}\\ \vdots\\ \widehat{\theta}_1^{s_{M_1^1}}\end{pmatrix}, \qquad (0.23)$$

де x_i, $i=\overline{1,M_1^1}$ – змінні (відповідні M_1^1 змінним виду $b_{i_1\ldots i_t}^{j_1\cdots j_t}$).

Нехай оцінки $\widehat{\theta}_1^{s_l}\ \ l=\overline{1,\grave{I}\ _1^1}$, заданої статистически значимою ймовірністю p оценивают $\theta_1^{S_l}$ з похибкаю, по модулю не превышающей чисел $\Delta_1^{S_l}\ l=\overline{1,M_1^1}$. Тодізймовірністю p максимальна похибка знаходження точних значень M_1^1 відповідних коефіцієнтів виду $b_{i_1\ \ldots i_t}^{j_1\cdots j_t}$ має вид

$$\max_{j=\overline{1,\bar{I}\ _1^{-1}}}\left\{\max\left(\sum_l{}^{(+)}a_{jl}^{-1}\Delta_1^{s_l},\sum_l{}^{(-)}\left|a_{jl}^{-1}\right|\Delta_1^{s_l}\right)\right\}, \tag{0.24}$$

де $\sum_l{}^{(+)}a_{jl}^{-1}\Delta_1^{s_l}$ береться по всім $l=\overline{1,M_1^1}$, для яких $a_{jl}^{-1}\geq 0;\ \ \sum_l{}^{(-)}a_{jl}^{-1}\Delta_1^{s_l}$ береться по всім $l=\overline{1,M_1^1}$, для яких $a_{jl}^{-1}<0;\ a_{jl}^{-1}-\ jl$-й елемент матриці A^{-1}.

Як вказувалося вище, передбачається, що $x_i^s,\ i=\overline{2,n},\ s=\overline{1,M_1^1}$ обрані так, що матриця A^{-1} існує.

Аналогічно будуються всі інші системи лінійних рівнянь (правими частинами яких є колонки $\left(\widehat{\theta}_l^{s_1},\ldots\widehat{\theta}_l^{s_{M_l^1}}\right)^{\mathsf{T}},\ l=\overline{2,n_1}$) для знаходження всіх інших коефіцієнтів $b_{i_1\ \ldots i_t}^{j_1\cdots j_t}$ із виразу (0.8). Аналогічно будуються всі оцінки виду (0.24).

Процедури знаходження всіх невідомих коефіцієнтів $b_{i_1\ \ldots i_t}^{j_1\cdots j_t}$ із виразів (0.9) для $l=\overline{2,n_1}$ повністю повторюютьпроцедуру, викладену для виразу (0.8).

Оцінка константи в виразі (0.6) може бути получена як середнє арифметичне по всім проведенным испытаниям разностей $y_i-\left(\hat{y}\left(\overline{x_i}\right)-\theta_0\right)$, де y_i – значеннявыходній змінної моделі, коли на вхід подается векторное значення $\overline{x_i}$, а выражение $\hat{y}\left(\overline{x_i}\right)-\theta_0$ – це значеннявиразу (0.7) для x_i, із которого исключен коефіцієнт θ_0 ізамість коефіцієнтів $b_{i_1\ \ldots i_t}^{j_1\cdots j_t}$ подставлены Отримані їх оцінки.

Якщо верхня оцінка σ_E^2 невідома, то її можна ефективно оцінити як середнє арифметичне оцінок σ_E^2 (0.15) по всім одновимірним регресіям.

АЛГОРИТМ РОЗВЯЗУВАННЯ ЗАДАЧІ

1. Рекомендації по проведению активного експерименту для одновимірного поліноміального регресійного аналізу

Проблема ефективної побудови поліноміальної регресії по даним із шумом часто зустрічається на практиці. Значну роль відіграє побудова відповідність побудованої регресійної моделі та дійсної моделі. В цьому розділі приводяться рекомендації відповідно зони проведення активного експеримента, по результатам якого буде проводитись регресійний аналіз. Ці рекомендації відносяться лише до тих випадків, коли дослідник може керувати проведенням експерименту і вибирати діапазон змін значень детермінованого аргументу. Розгляжується одновимірний поліноміальний аналіз з використанням ортогональних поліномів Фойсайта. Критеріем покращення регресійної моделі є зменшення дисперсій оцінок коефіцієнтів в регресійному поліномі. Приводяться прості формули перерахунку ортогональних нормованих поліномів за пропорційною зміною діапазона значень детермінованого аргументу. Далі приводяться аналітичний доказ того, що масштабування інтервалу після проведення ексеременту, не вплине на якість регресії. Експериментально буде показано, що повторювані експерименти несуттєво впливають на якість оцінки лінії регресії. Теоретично буде обгрунтовано при цьому значне спрощення обчислення коефіцієнтів лінії регресії.

Вибір зони проведення активного експерименту

Можна виділити шість умовних зон проведення експерименту:

1) Початок інтервалу більше нуля, малий інтервал. Наприклад, [10; 20].

2) Початок інтервалу більше нуля, інтервал великий. Наприклад, [10; 1010].

3) Початок інтервалу в точці нуль, інтервал малий. Наприклад, [0; 10].

4) Початок інтервалу в точці нуль, інтервал великий. Наприклад, [0; 1000].

5) Інтервал симетричний відносно точки нуль, інтервал малий. Наприклад, [–5; 5].

6) Інтервал симетричний відносно точки нуль, інтервал великий. *σ2*

В таблицях 3.1 – 3.3 приведені дисперсії оцінок коефіцієнтів регресійного поліному, розраховані по формулам, проведеним в розділі 6 [12].

×σ2

Інтервал	0	1	2	3	4	5
[10; 20]	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*
[10; 1010]	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*
[0; 10]	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*
[0; 1000]	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*
[–5;5]	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*
[–500; 500]	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*	*σ2*

$\times\sigma 2$

Інтервал	0	1	2	3	4	5
[10; 20]	$3{,}3\cdot10^{5}$	$4\cdot10^{4}$	$7{,}7\cdot10^{2}$	3,5	$4\cdot10^{-3}$	$7\cdot10^{-7}$
[10; 1010]	$6{,}1\cdot10^{-1}$	$2{,}1\cdot10^{-4}$	$7{,}1\cdot10^{-9}$	$4{,}2\cdot10^{-14}$	$4{,}7\cdot10^{-20}$	$7\cdot10^{-27}$
[0; 10]	$4{,}3\cdot10^{-1}$	1,7	$6{,}1\cdot10^{-1}$	$3{,}8\cdot10^{-2}$	$4{,}5\cdot10^{-4}$	$7\cdot10^{-7}$
[0; 1000]	$4{,}3\cdot10^{-1}$	$1{,}7\cdot10^{-4}$	$6{,}1\cdot10^{-9}$	$3{,}8\cdot10^{-14}$	$4{,}5\cdot10^{-20}$	$7\cdot10^{-27}$
[–5;5]	$3{,}5\cdot10^{-2}$	$2{,}3\cdot10^{-2}$	$2{,}2\cdot10^{-3}$	$5{,}7\cdot10^{-4}$	$4{,}5\cdot10^{-6}$	$7\cdot10^{-7}$
[–500; 500]	$3{,}5\cdot10^{-2}$	$2{,}3\cdot10^{-6}$ ∫ 2	$2{,}2\cdot10^{-11}$	$5{,}7\cdot10^{-16}$	$4{,}5\cdot10^{-22}$	$7\cdot10^{-27}$

$\times\sigma 2$

Інтервал	0	1	2	3	4	5
[10; 20]	$3{,}3\cdot10^{5}$	$4\cdot10^{4}$	$7{,}7\cdot10^{2}$	3,5	$4\cdot10^{-3}$	$7\cdot10^{-7}$
[10; 1010]	$6{,}1\cdot10^{-1}$	$2{,}1\cdot10^{-4}$	$7{,}1\cdot10^{-9}$	$4{,}2\cdot10^{-14}$	$4{,}7\cdot10^{-20}$	$7\cdot10^{-27}$
[0; 10]	$4{,}3\cdot10^{-1}$	1,7	$6{,}1\cdot10^{-1}$	$3{,}8\cdot10^{-2}$	$4{,}5\cdot10^{-4}$	$7\cdot10^{-7}$
[0; 1000]	$4{,}3\cdot10^{-1}$	$1{,}7\cdot10^{-4}$	$6{,}1\cdot10^{-9}$	$3{,}8\cdot10^{-14}$	$4{,}5\cdot10^{-20}$	$7\cdot10^{-27}$
[–5;5]	$3{,}5\cdot10^{-2}$	$2{,}3\cdot10^{-2}$	$2{,}2\cdot10^{-3}$	$5{,}7\cdot10^{-4}$	$4{,}5\cdot10^{-6}$	$7\cdot10^{-7}$
[–500; 500]	$3{,}5\cdot10^{-2}$	$2{,}3\cdot10^{-6}$	$2{,}2\cdot10^{-11}$	$5{,}7\cdot10^{-16}$	$4{,}5\cdot10^{-22}$	$7\cdot10^{-27}$

Як видно із таблиць 1-3, чим ширше інтервал, тем менше дисперсії оцінок коефіцієнтів. Також, дисперсії зменшуються, коли інтервал експерименту симетричний точці нуль. Якщо неможливо приймати відємні значення , то рекомендується починати інтервал в точці нуль. Інтервал досить сильно впливає на дисперсії оцінок коефіцієнтів, навіть більше ніж зміна кількості вхідних точок. Дисперсії оцінок коефіцієнтів можуть бути перераховані при масштабуванні по формулі (32):

$$D\hat{\theta}_j^z = \sigma^2 \sum_{l=r}^{j} (\frac{1}{k^j} q_{lj}^x)^2 = (\frac{1}{k^j})^2 D\hat{\theta}_j^x$$

З врахуванням того, що на практиці кількість вхідних точокі необхідна ступінь регресійного полінома можуть бути достатньо великими числами, це призводить до виконання значної кількості розрахунків. Перерахунок коефіцієнтів є достатньо простим процесом і не потребує значних компютерих потужностей. Розглянемо прерахунок дисперсій на прикладі (табл. 4).

Табл. 4. Приклад перерахутнку дисперсій оцінок коефіцієнтів (1000 вхідних точок)

	[-5;5]	[-500;500] $k = 100$
0	$3{,}5 \times 10^{-3} \sigma^2$	$\left(\frac{1}{100^0}\right)^2 \times 3{,}5 \times 10^{-3} \sigma^2 = 3{,}5 \times 10^{-3} \sigma^2$
1	$2{,}3 \times 10^{-3} \sigma^2$	$\left(\frac{1}{100^1}\right)^2 \times 2{,}3 \times 10^{-3} \sigma^2 = 2{,}3 \times 10^{-7} \sigma^2$
2	$2{,}2 \times 10^{-4} \sigma^2$	$\left(\frac{1}{100^2}\right)^2 \times 2{,}2 \times 10^{-4} \sigma^2 = 2{,}2 \times 10^{-12} \sigma^2$
3	$5{,}7 \times 10^{-5} \sigma^2$	$\left(\frac{1}{100^3}\right)^2 \times 5{,}7 \times 10^{-5} \sigma^2 = 5{,}7 \times 10^{-17} \sigma^2$
4	$4{,}4 \times 10^{-7} \sigma^2$	$\left(\frac{1}{100^4}\right)^2 \times 4{,}4 \times 10^{-7} \sigma^2 = 4{,}4 \times 10^{-23} \sigma^2$
5	$7 \times 10^{-8} \sigma^2$	$\left(\frac{1}{100^5}\right)^2 \times 7 \times 10^{-8} \sigma^2 = 7 \times 10^{-28} \sigma^2$

Таким чином, можна заздалегідь побачити ефективність регресійного аналізу, і підібрати необхідні досліднику діапазони. Аналогічно, по формулі (16) можна перерахувати значення коефіцієнтів нормованих ортогональних поліномів Форсайта: що потребує значно менше обчислень, ніж розразунок цих значень заново по формулам (7)-(10).

$$q_{jl}^z = \frac{1}{k^l} q_{jl}^x, \forall j = \overline{0, r}, \forall l = \overline{0, j}$$

Однако, проведення масштабування уже отриманих вхідних даних не приведе до покращення результату, далі це буде доведено.

1.1.Масштабування вхідних даних

Як показали експерименти, при масштабуванні уже наявних вхідних даних , якість регресійного полінома не змінюєіться. Тобто, недивлячись на те, що змінюються оцінки коефіцієнтіви їх дисперсії, значення регресійного полінома при заданному аргументі залишиться таким же.

Для доведення того, що при масштабуванні якість вхідних даних регресійного полінома не змінюється, будемо розглядати коефіцієнти нормованих ортогональних поліномів Форсайта окремо. Тоді із (7), (11) і (12) отримаємо вирази (14) - (20).

$$q_{00} = \frac{1}{\sqrt{n}} \tag{1}$$

$$q_{10} = -\frac{\bar{x}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{2}$$

$$q_{11} = \frac{1}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{3}$$

$$q_{j0} = \frac{\alpha q_{j-1,0} - \beta q_{j-2,0}}{\lambda} \tag{4}$$

$$q_{jl} = \frac{q_{j-1,l-1} - \alpha q_{j-1,l} - \beta q_{j-2,l}}{\lambda}, l = \overline{1, j-2} \tag{5}$$

$$q_{j,j-1} = \frac{q_{j-1,j-2} - \alpha q_{j-1,j-1}}{\lambda} \tag{6}$$

$$q_{j,j} = \frac{q_{j-1,j-1}}{\lambda} \tag{7}$$

Формули (14) - (20) виконуються $\forall j = \overline{2,r}$.

Нехай, було виконано масштабування вихідних даних наступним чином $z = kx$, де k – деякий коефіцієнт масштабування , який є дійсним числом.

Позначемо коефіцієнти при початковому масштабуванні з допомогою q^x, а при новому – q^z. Покажемо, що звязок коефіцієнтів нормованих ортогональних поліномів при різному масштабуванні відповідає формулі (21).

$$q_{jl}^z = \frac{1}{k^l} q_{jl}^x, \forall j = \overline{0,r}, \forall l = \overline{0,j} \qquad (8)$$

Розглянемо звязок між середім значенням аргументів початкового і нового масштабування , оскільки воно використовується в формулах (15)і(16).

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n} kx_i = k\bar{x} \qquad (9)$$

Розглянемо коефіцієнти перших двох нормованих ортогональних поліномів Форсайта

$$q_{00}^z = \frac{1}{\sqrt{n}} = q_{00}^x \qquad (10)$$

$$q_{10}^z = \frac{k\bar{x}}{\sqrt{\sum_{i=1}^{n}(kx_i - k\bar{x})^2}} = q_{10}^x \qquad (11)$$

$$q_{11}^z = \frac{1}{\sqrt{\sum_{i=1}^{n}(kx_i - k\bar{x})^2}} = \frac{1}{k} q_{11}^x \qquad (12)$$

Отже, для коефіцієнтів перших двох ортогональних поліномів виконується формула (21). Розглянемо значення ортогонального полінома при заданному аргументі, для яких виконується формула (21).

$$\theta_j^z(z_i) = q_{j0}^x + \cdots + \frac{1}{k^j} q_{jj}^x k^j x_i = \theta_j^x(x_i) \qquad (13)$$

Розглянемо звязок коефіцієнтів α, β, λ при різному масштабуванні.

$$\alpha^z = \sum_{i=1}^{n} kx_i Q_{j-1}^2(x_i) = k\alpha^x \qquad (14)$$

$$\beta^z = \sum_{i=1}^{n} kx_i \theta_{j-1}(x_i)\theta_{j-2}(x_i) = k\beta^x \qquad (15)$$

$$\lambda^z = \sqrt{\sum_{i=1}^{n} (kx_i\theta^x_{j-1}(x_i) - k\alpha\theta^x_{j-1}(x_i) - k\beta\theta^x_{j-2}(x_i))^2} = k\lambda^x \qquad (16)$$

Розглянемо звязок між коефіцієнтами поліномів $j = \overline{2,r}$ при різному масштабуванні.

$$q^z_{j0} = \frac{k\alpha^x q^x_{j-1,0} - k\beta^x q^x_{j-2,0}}{k\lambda^x} = q^x_{j0} \qquad (17)$$

$$q^z_{jl} = \frac{\frac{1}{k^{l-1}} q^x_{j-1,l-1} - k\alpha^x \frac{1}{k^l} q^x_{j-1,l} - k\beta^x \frac{1}{k^l} q^x_{j-2,l}}{k\lambda^x} = \frac{1}{k^l} q^x_{jl}, l = \overline{1, j-2} \qquad (18)$$

$$q^z_{j,j-1} = \frac{\frac{1}{k^{j-2}} q^x_{j-1,j-2} - k\alpha^x \frac{1}{k^{j-1}} q^x_{j-1,j-1}}{k\lambda^x} = \frac{1}{k^{j-1}} q^x_{j,j-1} \qquad (19)$$

$$q^z_{j,j} = \frac{\frac{1}{k^{j-1}} q^x_{j-1,j-1}}{k\lambda^x} = \frac{1}{k^j} q^x_{j,j} \qquad (20)$$

Отже, для ортогональних поліномів Форсайта формула (21) справедлива, при масштабуванні $z = kx$.

Розглянемо зв'язок оцінок вагових коефіцієнтів при різних масштабах.

$$\hat{w}^z_j = \sum_{i=1}^{n} y_i Q^x_j(x_i) = \hat{w}^x_j \qquad (21)$$

Виходячи із формул (21) і (34), зв'язок між оцінками коефіцієнтів регресійного полінома при різному масштабуванні буде наступним.

$$\hat{Q}^z_j = \hat{w}_r \frac{1}{k^j} q^x_{rj} + \cdots + \hat{w}_j \frac{1}{k^j} q^x_{jj} = \frac{1}{k^j} \hat{Q}^x_j \qquad (22)$$

Виходячи із (35), звязок між значеннями регресійного полінома при заданому аргументі і різному масштабуванні буде наступним.

$$f^z(z_i) = \sum_{j=0}^{r} \frac{1}{k^j} \hat{Q}^x_j k^j x_i = f^x(x_i) \qquad (23)$$

Отже, значення регресійного полінома, при різному масштабуванні наявних вхідних даних залишиться одинаковым, хоча вид самого регресійного полінома змінюється согласно формулі (35).

Розглянемо звязок між дисперсіями оцінок коефіцієнтів регресійного полінома при різному масштабуванні.

$$D\hat{Q}_j^z = \sigma^2 \sum_{l=r}^{j} (\frac{1}{k^j} q_{lj}^x)^2 = (\frac{1}{k^j})^2 D\hat{Q}_j^x \qquad (24)$$

Отже, дисперсії оцінок коефіцієнтів регресійного полінома змінюються по формулі (37).

4. Повторювані експерименти

В умовах задач, що зустрічаются на практиці досліднику потрібно проводити велику кількість експериментів на невеликому інтервалі значень аргументів. При рівномірному розподілі значень аргументів виникають деякі незручності, звязані з тим, що кожний наступний експеримент потрібно проводити змінив значення аргументу на деяке невелике число Δx, що складно на практиці, при дослідженні фізичних процесів. В цьому випадку на вхід зручно подавати повторювану послідовність

$$x_1, \cdots, x_{r+p}, x_1, \cdots, x_{r+p}, \cdots,$$

де $p \geq 1$, r – ступінь одновимірного полінома, який задається надлишковим описом.

Доведемо, що оцінки коефіцієнтів $\hat{\theta}_j, j = \overline{0,r}$ не змінюються при усередненні результатів експериментів, при виконанні одновимірного поліноміального регресійного аналізу методом із [1]. Введемо позначення :

$$X = (x_1, x_2, ..., x_n)$$

$$X' = (x_{11}, x_{21}, ...x_{l1}, x_{12}, x_{22}, ..., x_{l2}, ..., x_{1n}, x_{2n}, ..., x_{ln})$$

В X' знаходиться l копій X, тобто. $x_{ki} = x_i; \forall k = \overline{1,l}; \forall i = \overline{1,n}$.

$$Y' = (y_{11}, y_{21}, \ldots y_{l1}, \ldots, y_{1n}, y_{2n}, \ldots, y_{ln})$$

$$Y = (\frac{\sum_{k=1}^{l} y_{k1}}{l}, \frac{\sum_{k=1}^{l} y_{k2}}{l}, \ldots, \frac{\sum_{k=1}^{l} y_{kn}}{l})$$

Іншими словами, замість двух векторов X′ і Y′, розмірністю $n \times l$ кожний, ми подаемо X, Y, розмір яких – n, де n – кількість кожного із різних значень аргументу.

$Q_j, j = \overline{0,r}$ – ортогональні поліноми, побудовані на виборці X

$Q_j', j = \overline{0,r}$ – ортогональні поліноми, побудовані на виборці X′

Методом метематичної індукції покажемо, що

$$Q_j'(x) = \frac{Q_j(x)}{\sqrt{l}}, j = \overline{0,r} \qquad (38)$$

При j = 0 маємо:

$$Q_0'(x) = \frac{1}{\sqrt{n \times l}} = \frac{Q_0(x)}{\sqrt{l}}$$

При j = 1 маємо:

$$Q_1'(x) = \frac{x - \overline{X}'}{\sqrt{\sum_{i=1}^{n \times l} (x_i - \overline{X}')^2}} =$$

$$= \left[\begin{array}{l} \overline{X} = \overline{X}'; \\ \sum_{i=1}^{n \times l} (x_i - \overline{X}')^2 = \sum_{k=1}^{l} \sum_{i=1}^{n} (x_i - \overline{X}')^2 = l \times \sum_{i=1}^{n} (x_i - \overline{X}')^2 \end{array} \right] =$$

$$= -\frac{(x - \overline{X}')}{\sqrt{l \times \sum_{i=1}^{n} (x_i - \overline{X}')^2}} = \frac{(x - \overline{X}')}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{X}')^2}} \times \frac{1}{\sqrt{l}} = \frac{Q_1(x)}{\sqrt{l}}$$

Нехай (3) виконується для $Q_{j-1}'(x)$ и $Q_{j-2}'(x)$, тоді:

$$Q_{j-1}'(x) = \frac{Q_{j-1}(x)}{\sqrt{l}}$$

$$Q_{j-2}'(x) = \frac{Q_{j-2}(x)}{\sqrt{l}}$$

Покажемо тепер, що (38) виконується для будь-якого $Q_j'(x)$.

a) Спочатку доведемо , що $\alpha' = \alpha$, (тобто. α не залежить від того, чи використовує ми X чи X'):

$$\alpha' = \sum_{i=1}^{n\times l} x_i Q'^2_{j-1}(x_i) = \sum_{i=1}^{n\times l} x_i \frac{Q_{j-1}{}^2(x_i)}{l} = \sum_{k=1}^{l}\sum_{i=1}^{n} x_i \frac{Q_{j-1}{}^2(x_i)}{l}$$

Так як кожний елемент суми по індексу k не залежить від k, тоді ми можемо замінити суму на добуток:

$$\sum_{k=1}^{l}\sum_{i=1}^{n} x_i \frac{Q_{j-1}{}^2(x_i)}{l} = l\times \sum_{i=1}^{n} x_i \frac{Q_{j-1}{}^2(x_i)}{l} = \sum_{i=1}^{n} x_i Q_{j-1}{}^2(x_i) = \alpha$$

b) Аналогічно покажемо, що $\beta' = \beta$:

$$\beta' = \sum_{i=1}^{n\times l} x_i Q'_{j-1}(x_i) Q'_{j-2}(x_i) = \sum_{k=1}^{l}\sum_{i=1}^{n} x_i \frac{Q_{j-1}(x_i)}{\sqrt{l}} \times \frac{Q_{j-2}(x_i)}{\sqrt{l}} =$$

$$= l\times \sum_{i=1}^{n} x_i \frac{Q_{j-1}(x_i) Q_{j-2}(x_i)}{l} = \sum_{i=1}^{n} x_i Q_{j-1}(x_i) Q_{j-2}(x_i) = \beta$$

c) Аналогічно, покажемо що $\lambda' = \lambda$:

$$\lambda' = \sqrt{\sum_{i=1}^{n\times l} (x_i Q'_{j-1}(x_i) - \lambda' Q'_{j-1}(x_i) - \beta' Q'_{j-2}(x_i))^2} =$$

$$= \sqrt{\sum_{k=1}^{l}\sum_{i=1}^{n} (\frac{x_i Q_{j-1}(x_i)}{\sqrt{l}} - \frac{\lambda Q_{j-1}(x_i)}{\sqrt{l}} - \frac{\beta Q_{j-2}(x_i)}{\sqrt{l}})^2} =$$

$$= \sqrt{l\times \sum_{i=1}^{n} \frac{(x_i Q_{j-1}(x_i) - \lambda Q_{j-1}(x_i) - \beta Q_{j-2}(x_i))^2}{l}} =$$

$$= \sqrt{\sum_{i=1}^{n} (x_i Q_{j-1}(x_i) - \lambda Q_{j-1}(x_i) - \beta Q_{j-2}(x_i))^2} = \lambda$$

Тепер підставивши значення $\lambda', \beta', \alpha'$, доведем що виконується (38):

$$Q'_j(x) = \frac{xQ'_{j-1} - \alpha' Q'_{j-1} - \beta' Q'_{j-2}(x)}{\lambda'}$$

$$\frac{x\frac{Q_{j-1}}{\sqrt{l}} - \alpha\frac{Q_{j-1}}{\sqrt{l}} - \beta\frac{Q_{j-2}(x)}{\sqrt{l}}}{\lambda} =$$

$$= \frac{xQ_{j-1} - \alpha Q_{j-1} - \beta Q_{j-2}(x)}{\lambda} \times \frac{1}{\sqrt{l}} = \frac{Q_j(x)}{\sqrt{l}}$$

Ми показали, що при переході від рішення задачі регресії на X' до рішення на X, ми повинні замінити все значення поліномів по формулі (38).

Покажемо, що оцінка регресійного полінома не змінюєіться, при переході на данні, з усередненням результатів експериментів:

Спочатку, знайдемо залежність значень оцінок вагових коефіцієнтів $\hat{w}_j, j=\overline{0,r}$, отриманих при усередненні результатів експериментів і оцінок отриманих на даних без усереднення:

$$\hat{\omega}'_j = \sum_{i=1}^{n}\sum_{k=1}^{l} y_{ki} Q'_j(x_{ki}) = \left[Q'_j(x_{k_1 i}) = Q'_j(x_{k_2 i}), \forall k_1, k_2 \in \{1,\ldots,l\} \right] =$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{l} y_{ki} Q'_j(x_i) = \sum_{i=1}^{n}\sum_{k=1}^{l} y_{ki} \frac{Q_j(x_i)}{\sqrt{l}} = [\times \frac{l}{l}] = \sum_{i=1}^{n} \left(\frac{\sum_{k=1}^{l} y_{ki}}{l} \right) \times l \times \frac{Q_j(x_i)}{\sqrt{l}} = \hat{\omega}_j \times \sqrt{l}$$

Отримали , що:

$$\hat{\omega}'_j = \omega_j \times \sqrt{l}, \forall j = \overline{0,r}$$

Тепер покажемо, що $\hat{\theta}_j' = \hat{\theta}_j, \forall j = \overline{0,r}$:

$$\hat{\theta}_j' = \hat{w}_r' q'_{rj} + \cdots + \hat{w}'_j q'_{jj} = \hat{w}_r \sqrt{l} \times \frac{q_{rj}}{\sqrt{l}} + \cdots + \hat{w}_j \sqrt{l} \times \frac{q_{jj}}{\sqrt{l}} = \hat{\theta}_j, \forall j = \overline{0,r}$$

Дисперсії оцінок:

$$D\hat{Q}'_j = \frac{D\hat{Q}_j}{l}$$

Таким чином ми отримали наступний результат: в дослідженнях з повторюваними серіями експериментів при однаковому значенні аргументів, замість всіх даних на вхід алгоритму регресії можна подавати набагато менше точок, усереднюючи значення y при однаковому x при цьому оцінки коефіцієнтів не змінюються. Це дає значних виграш в кількості обчислень, так як сам алгоритм знаходження коефіцієнтів тепер не залежить від кількості всіх експериментів. Кількість обчислень залежить від кількості серій експериментів l і точок в одній серії n. Очевидно, що для ефективного відновлення потрібно щоб $n \times l > r$.

Очевидно, що масштабування вхідних даних , приведене в п. 2.2 Також буде справедливо і для повторюваних експериментів.

5. Висновок

При проведенні активного експерименту з отриманням вхідних даних регресії, дисперсії оцінок коефіцієнтів регресійного полінома будуть залежати від інтервалу, на якому він проводиться. чим ближче центр інтервалу до точки нуль, тем меншими будуть дисперсії. Однак значного пониження дисперсій можна досягнути шляхом розширення інтервалу. Однак, при масштабуванні уже наявних вхідних даних , якість регресійного полінома не змінюється, хоча змінюється його коефіцієнти і їх дисперсії. Для спрощення проведення експериментів рекомендується проводити серії експериментів на невеликій множині значень аргументів. В цій моделі експериментів можна усереднювати результати експериментів в серіях з однаковим значенням аргументів, що дає виграш в кількості обчислень. Також варто згадати той факт, що коефіцієнти при великих степенях x відновлюються краще ніж коефіцієнти при малих степенях.

2. Метод підвищеної точності по знаходженню оцінок коефіцієнтів багатовимірної регресії з надлишковим описом

В [1] наведено метод побудови багатовимірної поліноміальної регресії по надлишковому опису в умовах активного експерименту. Алгоритм базується на зведенні багатовимірної регресії к одновимірної з допомогою Фіксації всіх змінних окрім однієї. Цей метод має той недолік, що в одновимірних регресіях є велика кількість членів з невеликими значеннями степеней при x. Як було показано вище, коефіцієнти таких поліномів погано відновлюються.

В даній роботі наведено модифікований алгоритм побудови багатовимірної поліноміальної регресії, що базується на примітках в [1].

Нехай, як і раніше, багатовимірна модель задається у вигляді:

$$y\left(\overline{x}\right)=\sum_{\forall\left(i_1,\ldots i_r\right)\in K}\ \sum_{\forall\left(j_1,\ldots j_r\right)\in K\left(i_1,\ldots i_r\right)} b_{i_1\ldots i_r}^{j_1\ldots j_r}\left(x_{i_1}\right)^{j_1}\cdot\left(x_{i_2}\right)^{j_2}\ldots\left(x_{i_r}\right)^{j_r}+E, \tag{39}$$

де $\overline{x}=\left(x_1\ldots x_n\right)^{*}$ – детермінований вектор вхідних змінних;

x_i – i-та компонента вектора $\overline{x}$;

$b_{i_1\ldots i_r}^{j_1\ldots j_r}$ – вектор невідомих коефіцієнтів моделі (4), розмірністю r;

j_l – натуральні числа;

i_l – натуральні індекси із множини $\{1,\ldots,n\}$;

E – випадкова величина з нульовим математичним очікуванням і обмеженою невідомою дисперсією σ_E^2 (які в одновимірному випадку, може бути відома верхня оцінка σ_E^2).

Модель (39) є надлишковою – можливо, деякі із коефіцієнтів $b_{i_1\ldots i_t}^{j_1\ldots j_t}$ рівні нулю.

Введемо поняття фіксації змінних. Фіксуючи значення деякої підмножини змінних $x_i=x_{i\phi}\,|\,i\in\Phi$, де $\Phi-$множина індексів зафіксованих змінних, змінюючи значення інших змінних при наступних експериментах (в даної фіксації) однаково $x_i=x_j\,\forall i,j\notin\Phi$ багатовимірна регресія перетворюється в одновимірну

$$\theta_0+\theta_1x+\theta_2x^2+\ldots+\theta_nx^n \tag{40}$$

Де n – максимальний ступінь x при даній фіксації змінних.

Оцінки коефіцієнтів $\theta_0,\theta_1,\ldots,\theta_n$ а також дисперсії оцінок коефіцієнтів $D\theta_1,D\theta_2,\ldots,D\theta_n$ можна знайти по результатам одновимірного регресійного аналізу з використанням ортогональних поліномів Форсайта.

Модель (40) дозволяє побудувати систему із $n+1$ лінійних рівнянь, що зв'язують числа $\hat{\theta}_0,\hat{\theta}_1,\ldots,\hat{\theta}_n$ с коефіцієнтами $b_{i_1\ldots i_r}^{j_1\ldots j_r}$ моделі (39).

В ліву частину *i*-того рівняння входять коефіцієнти, які знаходяться при x в i -той ступені а також фіксовані змінні , в правій частині– оцінки $\hat{\theta}_0, \hat{\theta}_1, ..., \hat{\theta}_n$.

Проводячи фіксації змінних і отримуючи системи рівнянь, ми можемо об'єднати системи і розв'язати їх методом найменших квадратів.

Розв'язуючи систему (3) методом найменших квадратів, ми гарантовано отримаємо оцінки коефіцієнтів, якщо в матриці A знайдеться r лінійно незалежних рядків, де r – кількість невідомих коефіцієнтів вихідної моделі (не включаючи вільний член). Оскільки, за побудовою матриця A не має комплексних чисел, для перевірки кількості лінійно незалежних рядків можна використовувати той факт, що [2]

$$rang(A^T A) = rang(A) \tag{4}$$

Якщо $\det(A^T A) \neq 0$, то $rang(A) = r$. Тоді матриця $A^T A$ є невиродженої , і розв'язок (3) є невиродженим і єдиним.

Якщо в системі (3) є r лінійно незалежних рядків ($rang(A) = r$), то ми можемо стверджувати , що метод найменших квадратів дасть розв'язок, який і є коефіцієнтами вихідної моделі. В іншому випадку рішення задачі немає.

Новий метод базується на тому, що дисперсії оцінок коефіцієнтів одновимірного регресійного аналізу менше для коефіцієнтів які відповідають великим степеням. Фіксуючи малу кількість змінних ми збільшуємо ступені x в відповідній одновимірної регресії, тому із $r!$ можливих фіксацій змінних обираємо фіксації з найменшою кількістю фіксованих змінних $k = \overline{0, r}$. При фіксації k_0 змінних, можна задати $C_r^{k_0}$ комбінацій фіксацій k_0 змінних.

Опишем алгоритм знаходження коефіцієнтів моделі (1):

Алгоритм 1.1

Спочатку маємо пусту систему рівнянь для знаходження оцінок коефіцієнтів $b_{i_1 \ldots i_t}^{j_1 \cdots j_t}$. В матричному виде:

$$Ab = \theta \tag{3}$$

A - матриця коефіцієнтів системи;

b - вектор невідомих коефіцієнтів $b_{i_1 \ldots i_t}^{j_1 \cdots j_t}$ моделі (1), розмірністю r;

θ - вектор коефіцієнтів одновимірних регресій.

Крок 1. Фіксуємо набор змінних $\Phi_i^j, i = \overline{0, C_r^{k_0}}, j = \overline{1, r}$

Крок 2. Проводимо одновимірний регресійний аналіз і знаходимо оцінки коефіцієнтів $\theta_0^{ij}, \theta_1^{ij}, \ldots, \theta_{n^{ij}}^{ij}$ і їх дисперсії, будуємо систему рівнянь для даної фіксації і доповнюємо загальну систему рівнянь A.

Крок 4. Перевіряємо $\det(A^T A) \neq 0$, Якщо умова виконується то переходимо к шагу 5, в іншому випадку переходимо к нової фіксації (Крок 1).

Якщо були зафіксовані все перебори змінних і в системі все ще немає r незалежних рядків, то рішення задачі немає.

Крок 5. Вирішуємо систему методом найменших квадратів і отримаємо оцінки $\hat{b}_{i_1 \ldots i_t}^{j_1 \cdots j_t}$

Примітки

1. Проблема точності вектора правих частин вирішується наступним чином. В [1] показано, що оцінки коефіцієнтів одновимірних регресій при низьких степенях мають велику дисперсію . Таким чином, при наборе лінійних рівнянь із одновимірних регресій, варто додавати в загальну систему не все рівняння , а тільки те, праві частини яких містять θ_i^j при степенях x не менше 2. Це гарантує високу точність вектора правих частин і точне знаходження оцінок невідомих коефіцієнтів багатовимірної регресії. Якщо при такому підході в системі не набирається r лінійно незалежних рядків, то в неї можна включити

рівняння , в правих частинах яких знаходиться θ_i^j при x першої ступені. В системі не повинно бути рівнянь, відповідних вільному члену в побудованих одновимірних регресіях, так як оцінка вільного члена одновимірної регресії має велику дисперсію і погіршить якість оцінки коефіцієнтів багатовимірної регресії. Вільний член вихідної моделі знаходиться з допомогою оцінок інших коефіцієнтів, як середнє арифметичне відхилень значень регресійної моделі без нього, від вхідних даних .

2. На практиці виникають ситуації , коли число лінійно незалежних рядків менше кількості невідомих коефіцієнтів. В такому випадку ми можемо найти частковий розв'язок, вибравши із отриманої системи нову, невироджену підсистему з меншою кількістю змінних. Тут зручно використання людино-машинної процедури, так як людину може побачити і виділити такі підсистеми. Знаходження деяких коефіцієнтів початкової моделі набагато спрощує її і дозволяє вирішити меншу задачу регресії будь-яким другим способом.

Приклад

Нехай вихідна надлишкова модель має вигляд (1):

$$y(\overline{x}) = a_1 x_2^2 x_3^2 x_4^2 x_5^3 + a_2 x_1 + a_3 x_1 x_3 + a_4 x_2 x_5 + a_5 x_4^2 + a_6 x_2^2 x_3^3 x_4 + a_7 x_5^3 +$$
$$+ a_8 x_2^2 x_3^2 x_4^2 x_5^2 + a_9 x_1^2 x_2 x_4 x_5 + a_{10} x_4^2 x_5^2 + a_{11} x_1 x_3^3 x_5 + a_{12} x_4^3 + a_{13} x_4^4 + a_{14} x_3^3 x_5^3 +$$
$$+ a_{15} x_1^2 x_2 x_4^2 x_5 + a_{16} x_1^4 x_2 x_4^3 + a_{17} x_1^4 x_2^4 + a_{18} x_3 x_4 + a_{19} x_2^2 x_3^2 x_4^2 + a_{20} x_2^2 x_3^2 x_4^2 x_5 + E$$

$$a_1 = 0, a_2 = 10, a_3 = 11, a_4 = 12, a_5 = 13, a_6 = 14\ a_7 = 15, a_8 = 0, a_9 = 16,$$
$$a_{10} = 17, a_{11} = 18, a_{12} = 19, a_{13} = 0, a_{14} = 21, a_{15} = 22, a_{16} = 23, a_{17} = 24,$$
$$a_{18} = 25, a_{19} = 26, a_{20} = 0$$

E - випадкова величина з нульовим математичним очікуванням $ME = 0$ і обмеженою дисперсією $\sigma^2 = 50$.

Фіксуємо 0 змінних, тобто. змінюємо всі аргументи однаково.

При зміні всіх аргументів однаково $x_1 = x_2 = ... x_r = x$

багатовимірна регресія перетворюється в одновимірну:

$$y(\bar{x}) = a_2 x + (a_3 + a_4 + a_5 + a_{18})x^2 + (a_7 + a_{12})x^3 + (a_{10} + a_{13})x^4 +$$
$$+(a_9 + a_{11})x^5 + (a_6 + a_{14} + a_{15} + a_{19})x^6 + a_{20}x^7 + (a_8 + a_{16} + a_{17})x^8 + a_1 x^9 + E$$

Виконуємо заміну

$a_2 = \theta_1^{0,1}$

$a_3 + a_4 + a_5 + a_{18} = \theta_2^{0,1}$

$a_7 + a_{12} = \theta_3^{0,1}$

$a_{10} + a_{13} = \theta_4^{0,1}$

$a_9 + a_{11} = \theta_5^{0,1}$

$a_6 + a_{14} + a_{15} + a_{19} = \theta_6^{0,1}$

$a_{20} = \theta_7^{0,1}$

$a_8 + a_{16} + a_{17} = \theta_8^{0,1}$

$a_1 = \theta_9^{0,1}$

Проводимо експерименти, в яких ми змінюємо x в діапазоні [-100, 100] з кроком 2. З допомогою одновимірного регресійного аналізу знаходимо оцінки коефіцієнтів при x^j:

Коефіцієнт	$\hat{\theta}_1^{0,1}$	$\hat{\theta}_2^{0,1}$	$\hat{\theta}_3^{0,1}$	$\hat{\theta}_4^{0,1}$	$\hat{\theta}_5^{0,1}$	$\hat{\theta}_6^{0,1}$	$\hat{\theta}_7^{0,1}$	$\hat{\theta}_8^{0,1}$	$\hat{\theta}_9^{0,1}$
Оцінки	6.99	58.68	34.01	17	34	83	0	47	0
Дисперсія оцінок	$4.7\times 10^{-4}\sigma^2$	$4.37\times 10^{-7}\sigma^2$	$5.74\times 10^{-10}\sigma^2$	$9.65\times 10^{-14}\sigma^2$	$6.99\times 10^{-17}\sigma^2$	$2.51\times 10^{-21}\sigma^2$	$1.27\times 10^{-24}\sigma^2$	$6.64\times 10^{-30}\sigma^2$	$2.63\times 10^{-33}\sigma^2$

Отримані оцінки дозволяють побудувати 9 рівнянь, які ми додаємо в загальну систему A:

$$\begin{cases} a_2 = \hat{\theta}_1^{0,1} \\ a_3 + a_4 + a_5 + a_{18} = \hat{\theta}_2^{0,1} \\ a_7 + a_{12} = \hat{\theta}_3^{0,1} \\ a_{10} + a_{13} = \hat{\theta}_4^{0,1} \\ a_9 + a_{11} = \hat{\theta}_5^{0,1} \\ a_6 + a_{14} + a_{15} + a_{19} = \hat{\theta}_6^{0,1} \\ a_{20} = \hat{\theta}_7^{0,1} \\ a_8 + a_{16} + a_{17} = \hat{\theta}_8^{0,1} \\ a_1 = \hat{\theta}_9^{0,1} \end{cases}$$

Очевидно, що $rang(A) < r$, переходимо до наступної фіксації.

Фіксуємо змінну $x_5 = x_{5\phi} = 80.386$, інші змінні нехай змінюються однаково $x_1 = x_2 = x_3 = x_4 = x$. Тоді багатовимірна регресія перетворюється в одновимірну із згрупованими коефіцієнтами:

$$y(\overline{x}) = a_7 x_{5\phi}^3 + (a_2 + a_4 x_{5\phi})x + (a_3 + a_5 + a_{10} x_{5\phi}^2 + a_{18})x^2 + (a_{12} + a_{14} x_{5\phi}^3)x^3 +$$
$$+(a_9 x_{5\phi} + a_{11} x_{5\phi} + a_{13})x^4 + a_{15} x_{5\phi} x^5 + (a_1 x_{5\phi}^3 + a_6 + a_8 x_{5\phi}^2 + a_{19} + a_{20} x_{5\phi})x^6 + (a_{16} + a_{17})x^8 + E$$

Проводимо заміну:

$$a_2 + a_4 x_{5\phi} = \theta_1^{1,1}$$
$$a_3 + a_5 + a_{10} x_{5\phi}^2 + a_{18} = \theta_2^{1,1}$$
$$a_{12} + a_{14} x_{5\phi}^3 = \theta_3^{1,1}$$
$$a_9 x_{5\phi} + a_{11} x_{5\phi} + a_{13} = \theta_4^{1,1}$$
$$a_{15} x_{5\phi} = \theta_5^{1,1}$$
$$a_1 x_{5\phi}^3 + a_6 + a_8 x_{5\phi}^2 + a_{19} + a_{20} x_{5\phi} = \theta_6^{1,1}$$
$$a_{16} + a_{17} = \theta_8^{1,1}$$

Проводимо експерименти, в яких ми змінюємо x в діапазоні [-100, 100] з кроком 2. З допомогою одновимірного регресійного аналізу знаходимо оцінки коефіцієнтівпри x^j:

Коефіцієнт	$\hat{\theta}_1^{1,1}$	$\hat{\theta}_2^{1,1}$	$\hat{\theta}_3^{1,1}$	$\hat{\theta}_4^{1}$	$\hat{\theta}_5^{1,1}$	$\hat{\theta}$	$\hat{\theta}$
Оцінки	9 89.72	1 09899.5	1090 8452.41	2 733.13	17 68.49	4 0	4 7
Дисперсія оцінок	2 1	4.3 10^{-}	$1.2\times$ $10^{-10}\sigma^2$	9. 10	4.84 10^{-18}	2 1	6 1

Доповнивши загальну систему новими рівняннями, отримаємо:

$$
\begin{cases}
a_2 = \hat{\theta}_1^{0,1} \\
a_3 + a_4 + a_5 + a_{18} = \hat{\theta}_2^{0,1} \\
a_7 + a_{12} = \hat{\theta}_3^{0,1} \\
a_{10} + a_{13} = \hat{\theta}_4^{0,1} \\
a_9 + a_{11} = \hat{\theta}_5^{0,1} \\
a_6 + a_{14} + a_{15} + a_{19} = \hat{\theta}_6^{0,1} \\
a_{20} = \hat{\theta}_7^{0,1} \\
a_8 + a_{16} + a_{17} = \hat{\theta}_8^{0,1} \\
a_1 = \hat{\theta}_9^{0,1} \\
a_2 + a_4 x_{5\phi} = \hat{\theta}_1^{1,1} \\
a_3 + a_5 + a_{10} x_{5\phi}^2 + a_{18} = \hat{\theta}_2^{1,1} \\
a_{12} + a_{14} x_{5\phi}^3 = \hat{\theta}_3^{1,1} \\
a_9 x_{5\phi} + a_{11} x_{5\phi} + a_{13} = \hat{\theta}_4^{1,1} \\
a_{15} x_{5\phi} = \hat{\theta}_5^{1,1} \\
a_1 x_{5\phi}^3 + a_6 + a_8 x_{5\phi}^2 + a_{19} + a_{20} x_{5\phi} = \hat{\theta}_6^{1,1} \\
a_{16} + a_{17} = \hat{\theta}_8^{1,1}
\end{cases}
$$

Очевидно, що $rang(A) < r$, так як кількість рядків в системі в матричному вигляді менше r переходимо до наступної фіксації.

Фіксуємо змінну $x_4 = x_{4\phi} = -86.158$, інші змінні нехай змінюються однаково $x_1 = x_2 = x_3 = x_5 = x$. Тоді багатовимірна регресія перетворюється в одновимірну із згрупованими коефіцієнтами:

$$\overline{y(x)} = a_5 x_{4\phi}^2 + a_{12} x_{4\phi}^3 + a_{13} x_{4\phi}^4 + (a_2 + a_{18} x_4)x + (a_3 + a_4 + a_{10} x_{4\phi}^2)x^2 +$$
$$+ a_7 x^3 + (a_9 x_{4\phi} + a_{15} x_{4\phi}^2 + a_{19} x_{4\phi}^2)x^4 + (a_6 x_{4\phi} + a_{11} + a_{16} x_{4\phi}^3 + a_{20} x_{4\phi}^2)x^5 +$$
$$+ (a_8 x_{4\phi}^2 + a_{14})x^6 + a_1 x_{4\phi}^2 x^7 + a_{17} x^8$$

Проводимо заміну:

$$a_2 + a_{18} x_4 = \theta_1^{1,2}$$
$$a_3 + a_4 + a_{10} x_{4\phi}^2 = \theta_2^{1,2}$$
$$a_7 = \theta_3^{1,2}$$
$$a_9 x_{4\phi} + a_{15} x_{4\phi}^2 + a_{19} x_{4\phi}^2 = \theta_4^{1,2}$$
$$a_6 x_{4\phi} + a_{11} + a_{16} x_{4\phi}^3 + a_{20} x_{4\phi}^2 = \theta_5^{1,2}$$
$$a_8 x_{4\phi}^2 + a_{14} = \theta_6^{1,2}$$
$$a_1 x_{4\phi}^2 = \theta_7^{1,2}$$
$$a_{17} = \theta_8^{1,2}$$

Проводимо експерименти, в яких ми змінюємо x в діапазоні [-100, 100] з кроком 2. З допомогою одновимірного регресійного аналізу знаходимо оцінки коефіцієнтів при x^j:

Коефіцієнт	$\hat{\theta}_1^{1,2}$	$\hat{\theta}_2^{1,2}$	$\hat{\theta}_3^{1,2}$	$\hat{\theta}_4^{1,2}$	$\hat{\theta}_5^{1,2}$	$\hat{\theta}_6^{1,2}$	$\hat{\theta}_7^{1,2}$	$\hat{\theta}_8^{1,2}$
Оцінки	-2134.19	126216.4	14.99	354936.02	-14711311.32	21	0	24

Дисперсія оцінок	$2.47\times 10^{-4}\sigma^2$	$4.37\times 10^{-7}\sigma^2$	$1.2\times 10^{-10}\sigma^2$	$9.54\times 10^{-14}\sigma^2$	$4.84\times 10^{-18}\sigma^2$	$2.51\times 10^{-21}\sigma^2$	6.64×10^{-30}	$2.47\times 10^{-4}\sigma^2$

Доповнивши загальну систему новими рівняннями, отримаємо:

$$
\begin{cases}
a_2 = \hat{\theta}_1^{0,1} \\
a_3 + a_4 + a_5 + a_{18} = \hat{\theta}_2^{0,1} \\
a_7 + a_{12} = \hat{\theta}_3^{0,1} \\
a_{10} + a_{13} = \hat{\theta}_4^{0,1} \\
a_9 + a_{11} = \hat{\theta}_5^{0,1} \\
a_6 + a_{14} + a_{15} + a_{19} = \hat{\theta}_6^{0,1} \\
a_{20} = \hat{\theta}_7^{0,1} \\
a_8 + a_{16} + a_{17} = \hat{\theta}_8^{0,1} \\
a_1 = \hat{\theta}_9^{0,1} \\
a_2 + a_4 x_{5\phi} = \hat{\theta}_1^{1,1} \\
a_3 + a_5 + a_{10} x_{5\phi}^2 + a_{18} = \hat{\theta}_2^{1,1} \\
a_{12} + a_{14} x_{5\phi}^3 = \hat{\theta}_3^{1,1} \\
a_9 x_{5\phi} + a_{11} x_{5\phi} + a_{13} = \hat{\theta}_4^{1,1} \\
a_{15} x_{5\phi} = \hat{\theta}_5^{1,1} \\
a_1 x_{5\phi}^3 + a_6 + a_8 x_{5\phi}^2 + a_{19} + a_{20} x_{5\phi} = \hat{\theta}_6^{1,1} \\
a_{16} + a_{17} = \hat{\theta}_8^{1,1} \\
a_2 + a_{18} x_4 = \hat{\theta}_1^{1,2} \\
a_3 + a_4 + a_{10} x_{4\phi}^2 = \hat{\theta}_2^{1,2} \\
a_7 = \hat{\theta}_3^{1,2} \\
a_9 x_{4\phi} + a_{15} x_{4\phi}^2 + a_{19} x_{4\phi}^2 = \hat{\theta}_4^{1,2} \\
a_6 x_{4\phi} + a_{11} + a_{16} x_{4\phi}^3 + a_{20} x_{4\phi}^2 = \hat{\theta}_5^{1,2} \\
a_8 x_{4\phi}^2 + a_{14} = \hat{\theta}_6^{1,2} \\
a_1 x_{4\phi}^2 = \hat{\theta}_7^{1,2} \\
a_{17} = \hat{\theta}_8^{1,2}
\end{cases}
$$

Отримана система є невиродженої , так як $\det(A^T A) \neq 0$. Вирішуємо систему методом найменших квадратів. Результати приведені в таблице:

ефіцієнт	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
йсное ачение	0	0	1	12	13	14	15	0	16	17
одифікован і метод	-6.79×10^{-14}	6.99	9.12	12.23	12.49	13.99	14.99	-5.46×10^{-12}	16.93	$17+4.4\times 10^{-5}$
етод із [1]	-1.1×10^{-10}	-5.8×10^{-4}	18.59	0.21	-1.9×10^{-3}	$14+6.4\times 10^{-7}$	15.17	-4.57×10^{-9}	$16-1.0\times 10^{-3}$	17.06
ефіцієнт	a_{11}	a_{11}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	a_{18}	a_{19}	a_{20}
йсное ачение	18	9	0	21	22	23	24	25	26	0
одифікован і метод	17.07	19.02	0.001	$21-7.3\times 10^{-7}$	$22+3.6\times 10^{-8}$	$23+3.1\times 10^{-11}$	$24-1.2\times 10^{-10}$	24.85	26.01	5.91×10^{-10}
етод із [1]	$18-4.3\times 10^{-5}$	$19-4.4\times 10^{-5}$	3.2×10^{-7}	$21+1.9\times 10^{-7}$	$22+1.4\times 10^{-6}$	$23-6.2\times 10^{-15}$	$24-8.7\times 10^{-14}$	8.6	$26+2.1\times 10^{-5}$	5.98×10^{-7}

Таким чином, ми визначили що коефіцієнти a_1, a_8, a_{13}, a_{20} рівні нулю і дисперсії оцінок правих частин системи малі, ми можемо спростити надлишковий опис, виключив із нього відповідні члени.

Відновлена залежність:

$$y(\bar{x}) = 10.65x_1 + 11.09x_1x_3 + 12.14x_2x_5 + 13.38x_4^2 + 13.99x_2^2x_3^3x_4 + 14.99x_5^3 + 15.99x_1^2x_2x_4x_5 + 16.99x_4^2x_5^2$$
$$+17.99x_1x_3^3x_5 + 19x_4^3 + 21x_4^4 + 21.99x_1^2x_2x_4^2x_5$$
$$+22.99x_1^4x_2x_4^3 + 23.99x_1^4x_2^4 + 24.86x_3x_4 + 26x_2^2x_3^2x_4^2$$

ОПИС РОЗРОБЛЕНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

4 ОПИС РОЗРОБЛЕНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

4.1 Призначення програмного забезпечення

Розроблене програмне забезпечення призначене для автоматизації процесу планування роботи на Спринт шляхом розв'язування задачі розробленими алгоритмами. Програмне забезпечення має такий функціонал:

- зчитування вхідних даних задачі з обраного користувачем файлу;
- розв'язування задачі розробленими алгоритмами;
- виведення результатів, отриманих кожним із алгоритмів.

4.2 Засоби розробки

4.2.1 Вибір мови програмування

В якості основної мови програмування, на якій створюється ядро програми (алгоритм багатовимірної поліноміальної регресії) була обрана мова програмування Python [17,18,19], так як ця мова вважається стандартом у обробці даних та прототипуванні алгоритмів[20]. Плюсами є велика кількість математичних модулів, що можуть бути використаними під час розробки, та широка спільнота розробників, завдяки якій можна легко отримати відповіді на свої запитання по обробці даних.

Альтернативами могли би бути Matlab [21] або R [22]; Matlab був відкинутий, адже оскільки він є комерційним продуктом [23]. До того ж, модуль написаний на Matlab було б важче інтегрувати з іншими частинами системи.

R являється гарною альтернативою, але все ж не був обраний як основна мова. Python був більш доцільним, адже основний модуль повинен бути доступний через REST [24] API, що набагато простіше реалізувати на Python.

Пайтон (Python) — це потужна мова програмування, якою легко оволодіти. Вона має ефективні структури даних високого рівня та простий, але ефективний підхід до об'єктно-орієнтованого програмування. Елегантний синтаксис Пайтона, динамічна обробка типів, а також те, що це інтерпретована мова,

роблять його ідеальним для написання скриптів та швидкої розробки прикладних програм у багатьох галузях на більшості платформ.

Інтерпретатор мови Пайтон і багата стандартна бібліотека (як код-джерело, так і бінарні дистрибутиви для усіх головних операційних систем) можуть бути отримані з сайту Пайтона, і можуть вільно розповсюджуватися. Цей самий сайт має дистрибутиви та посилання на численні модулі, програми, утиліти та додаткову документацію.

Інтерпретатор мови Пайтон може бути легко розширений функціями та типами даних, розробленими на C чи C++ (або на іншій мові, яку можна викликати із C). Пайтон також зручний як мова сценаріїв що вбудовуються в прикладні програми, для додаткових налаштувань функціональності.

Переваги

Серед основних її переваг можна назвати такі:

– чистий синтаксис (для виділення блоків слід використовувати відступи);

– переносність програм (що властиве більшості інтерпретованих мов);

– стандартний дистрибутив має велику кількість корисних модулів (включно з модулем для розробки графічного інтерфейсу);

– можливість використання Python в діалоговому режимі (дуже корисне для експериментування та розв'язання простих задач);

– стандартний дистрибутив має просте, але разом із тим досить потужне середовище розробки, яке зветься IDLE [25] і яке написане на мові Python;

– зручний для розв'язання математичних проблем (має засоби роботи з комплексними числами, може оперувати з цілими числами довільної величини, у діалоговому режимі може використовуватися як потужний калькулятор).

Python має ефективні структури даних високого рівня та простий, але ефективний підхід до об'єктно-орієнтованого програмування. Елегантний

синтаксис Python, динамічна обробка типів, а також те, що це інтерпретована мова, роблять її ідеальною для написання скриптів та швидкої розробки прикладних програм у багатьох галузях на більшості платформ.

Інтерпретатор мови Python і багата стандартна бібліотека (як вихідні тексти, так і бінарні дистрибутиви для всіх основних операційних систем) можуть бути отримані з сайту Python www.python.org, і можуть вільно розповсюджуватися. Цей самий сайт має дистрибутиви та посилання на численні модулі, програми, утиліти та додаткову документацію. Інтерпретатор мови Python може бути розширений функціями та типами даних, розробленими на C чи C (або на іншій мові, яку можна викликати із C). Python також зручна як мова розширення для прикладних програм, що потребують подальшого налагодження.

Історія

Розробка мови Python була розпочата в кінці 1980-х років співробітником голландського інституту CWI Гвідо ван Россумом. Для розподіленої ОС Amoeba потрібна була розширювана скриптова мова, і Гвідо почав писати Python на дозвіллі, запозичивши деякі напрацювання для мови ABC (Гвідо брав участину у розробці цієї мови, орієнтованої на навчання програмуванню). У лютому 1991 року Гвідо опублікував вихідний текст в групі новин alt.sources. Мова почала вільно поширюватися через Інтернет, і сподобалася іншим програмістам. З 1991 року Python є цілком об'єктно- орієнтованим. Python також запозичив багато рис таких мов, як C, C++, Modula-3 і Icon, й окремі риси функціонального програмування з Ліспу.

Назва мови виникла зовсім не від виду плазунів. Автор назвав мову на честь популярного британського комедійного серіалу 70-х років «Повітряний цирк Монті Пайтона». Втім, все одно назву мови частіше асоціюють саме зі змією, ніж з фільмом — піктограми файлів в KDE або в Windows, і навіть емблема на сайті python.org зображують зміїну голову.

Наявність дружньої спільноти користувачів вважається, поряд з дизайнерською інтуїцією Гвідо, одним з головних факторів успіху Python. Розвиток мови відбувається згідно з чітко регламентованими процесами створення, обговорення, відбору та реалізації документів PEP (Python Enhancement Proposal) — пропозицій щодо розвитку Python.

3 грудня 2008 року, після тривалого тестування, вийшла перша версія Python 3000 (або Python 3.0, також використовується скорочена Py3k). У Python 3000 усунено багато недоліків архітектури з максимально можливим (але не повним) збереженням сумісності зі старими версіями. На сьогодні підтримуються обидві гілки розвитку (Python 3.2 і 2.7) [17].

Філософія

Розробники мови Python є прихильниками певної філософії програмування, яку називають «The Zen of Python» («Дзен Пайтона»)[26]. Її текст можна отримати у інтерпретаторі Python за допомогою команди import this (лише один раз за сесію). Автором цієї філософії вважається Тім Пейтерс [27].

Текст філософії:

– гарне краще, ніж потворне;

– явне краще, ніж неявне;

– просте краще, ніж складне;

– складне краще, ніж заплутане;

– плоске краще, ніж вкладене;

– розріджене краще, ніж щільне;

– легкість читання має значення;

– особливі випадки не настільки особливі, аби порушувати

– правила;

– при цьому практичність важливіше бездоганності;

– помилки ніколи не повинні замовчуватися;

– якщо не замовчуються явно;

– зустрівши двозначність, відкинь спокусу вгадати;

– має існувати один — і, бажано, тільки один — очевидний спосіб

– зробити це;

– хоча спочатку він може бути і не очевидним, якщо ви не

– голландець

– зараз краще, ніж ніколи;

– хоча ніколи, як правило, краще, ніж прямо зараз;

– якщо реалізацію важко пояснити — ідея погана;

– якщо реалізацію легко пояснити — ідея, можливо, добра;

– простори імен — чудова річ! Будемо робити їх побільше;

Сторонні бібліотеки

Numpy — розширення мови Python, що додає підтримку великих багатовимірних масивів і матриць, разом з великою бібліотекою високорівневих математичних функцій для операцій з цими масивами. Попередник Numpy, Numeric, був спочатку створений Jim Hugunin. Numpy — відкрите програмне забезпечення і має багато розробників.

Оскільки Python — інтерпретована мова, математичні алгоритми, часто працюють в ньому набагато повільніше ніж у компільованих мовах, таких як C або навіть Java. NumPy намагається вирішити цю проблему для великої кількості обчислювальних алгоритмів забезпечуючи підтримку багатовимірних масивів і безліч функцій і операторів для роботи з ними. Таким чином будь-який алгоритм який може бути виражений в основному як послідовність операцій над масивами і матрицями працює також швидко як еквівалентний код написаний на C.

NumPy можна розглядати як гарну вільну альтернативу MATLAB, оскільки мова програмування MATLAB зовні нагадує NumPy: обидві вони інтерпретовані, і обидві дозволяють користувачам писати швидкі програми поки більшість операцій проводяться над масивами або матрицями, а не над скалярами. Перевага MATLAB у великій кількості доступних додаткових тулбоксів, включаючи такі як пакет Simulink. Основні пакети, що доповнюють NumPy, це: SciPy — бібліотека, що додає більше MATLAB-подібної

функціональності; Matplotlib — пакет для створення графіки в стилі MATLAB. Внутрішньо як MATLAB, так і NumPy базується на бібліотеці LAPACK, призначеної для вирішення основних задач лінійної алгебри.

pandas — програмна бібліотека, написана для мови програмування Python для маніпулювання даними та їхнього аналізу. Вона, зокрема, пропонує структури даних та операції для маніпулювання чисельними таблицями та часовими рядами. pandas є вільним програмним забезпеченням, що випускається за трипунктовою ліцензією BSD[4]. Ця назва походить від терміну «панельні дані[en]» (англ. *panel data*), який в економетрії позначає багатовимірні структуровані набори даних.

- Об'єкт DataFrame із вбудованим індексуванням для маніпулювання даними.
- Інструменти для зчитування та записування даних між структурами даних у пам'яті та різними форматами файлів.
- Вирівнювання даних та вбудована підтримка пропущених даних.
- Переформатовування для отримання зведених наборів даних.
- Отримання зрізів за мітками, індексування з розширеними можливостями[5] та отримання піднаборів з великих наборів даних.
- Вставляння та вилучення стовпчиків у структурах даних.
- Рушій групування, що дозволяє робити з наборами даних операції розділення-зміни-об'єднання (англ. *split-apply-combine*).
- Злиття та з'єднання наборів даних.
- Ієрархічне індексування осей для роботи з даними високої вимірності в структурі даних нижчої вимірності.
- Функціональність для часових рядів: породження діапазонів дат та перетворення частоти, статистики рухливого вікна, лінійні регресії рухливого вікна, зсування дат та запізнювання.

Об'єктно-орієнтоване програмування [28]

Дизайн мови Python побудований навколо об'єктно-орієнтованої моделі програмування. Реалізація ООП в Python є елегантною, потужною та добре продуманою, але разом з тим, достатньо специфічною в порівнянні з іншими об'єктно-орієнтованими мовами.

Можливості та особливості:

– класи є одночасно об'єктами з усіма нижче наведеними можливостями;

– успадкування, в тому числі множинне;

– поліморфізм (всі функції віртуальні);

– інкапсуляція (два рівні — загальнодоступні та приховані методи і

– поля). Особливість — приховані члени доступні для використання та помічені як приховані лише особливими іменами;

– спеціальні методи, що керують життєвим циклом об'єкта: конструктори, деструктори, розподільники пам'яті;

– перевантаження операторів (усіх, крім is, '.', '=' і символьних логічних);

– властивості (імітація поля за допомогою функцій);

– управління доступу до полів (емуляція полів і методів, частковий доступ тощо);

– методи для управління найпоширенішими операціями (істинносне значення, len(), глибоке копіювання, серіалізація, ітерація по об'єкту);

– метапрограмування (управління створенням класів, тригери на створення класів, та ін);

– повна інтроспекція;

– класові та статичні методи, класові поля;

– класи, вкладені у функції та інші класи.

Функціональне програмування

Python підтримує парадигму функціонального програмування, зокрема:

– функція є об'єктом;

– функції вищих порядків;

– рекурсія;

– розвинена обробка списків (спискові вирази, операції над

– послідовностями, ітератори);

– аналог замикань (closures);

– часткове застосування функції;

– можливість реалізації інших засобів на самій мові (наприклад,

– каррінг).

4.3 Опис програмної реалізації

На рисунку 4.1 наведено схему структурну варіантів використання розробленого програмного забезпечення.

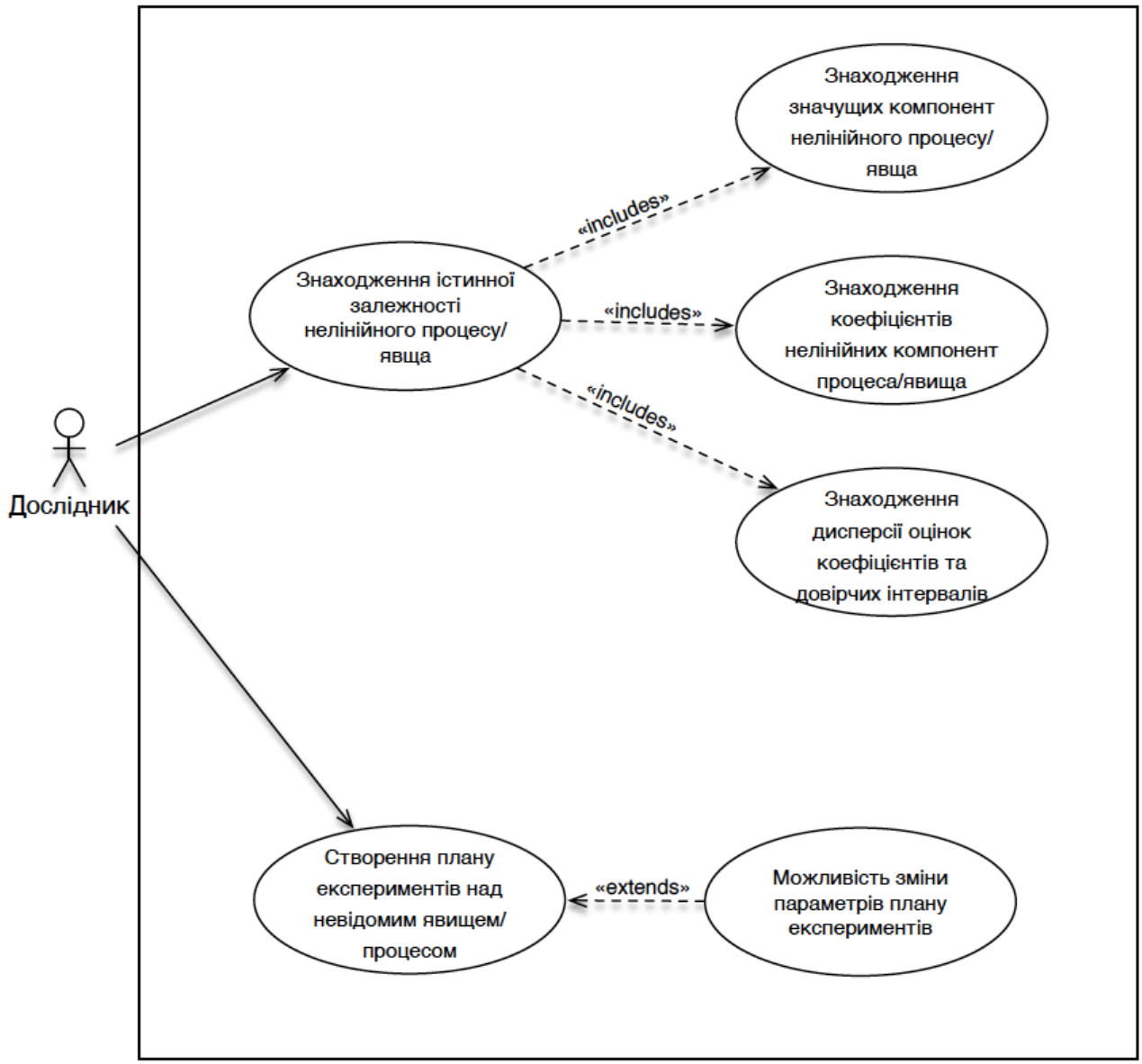

Рисунок 4.1 – Схема структурна варіантів використання

У таблиці 4.1 наведено опис класів модулю розв’язування підзадачі 1.

Таблиця 4.1 – Класи модулю розв’язування підзадачі розподілу обов’язкових завдань

Назва класу	**Функціональність**
MPRSolver	Клас, який містить в собі реалізацію алгоритму розв’язування задачі про рюкзак

Назва класу	Функціональність
Expression	Клас, який містить в собі реалізацію алгоритму розв’язування узагальненої задачі про призначення
RandGenerator	Клас, який представляє собою учасника команди, має номер
Term	Представляє собою завдання, яке має номер

На рисунку 4.2наведено схеми структурну класів модулю, в якому розміщено програмну реалізацію алгоритмів для розв’язування першої підзадачі з підрозділу 2.4.

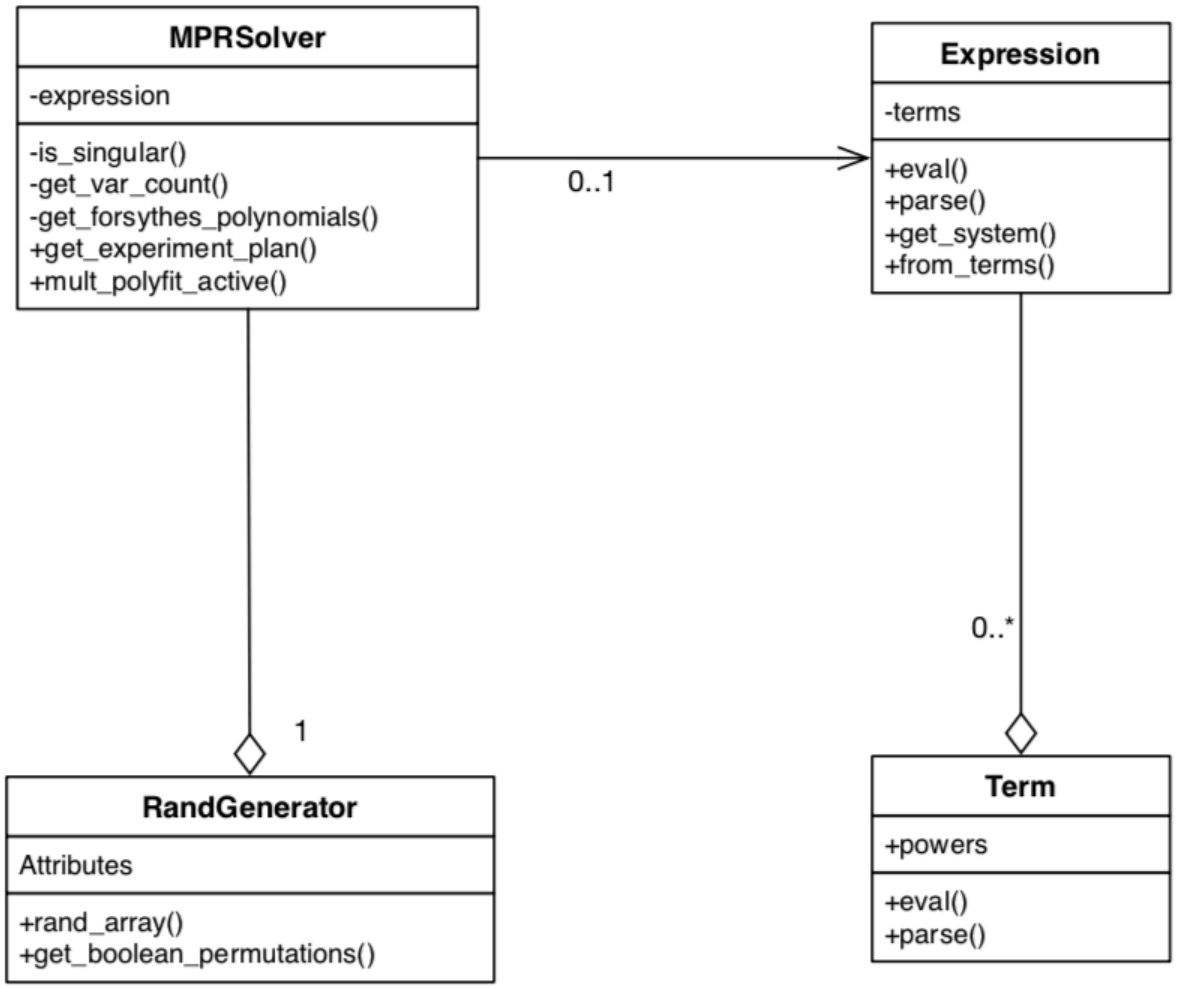

Рисунок 4.2 – Схема структурна класів модулю розв’язування підзадачі розподілу обов’язкових завдань

У таблиці 4.2 наведено опис класів модулю розв’язування підзадачі 2.

Таблиця 4.3 – Опис основних функцій модулів розв'язування першої та другої підзадач

Таблиця 3.1 – Специфікація функцій

Функція	**Опис функції**
get_boolean_permutation	Отримати всі перестановки заданих чисел
from_terms	Розібрати стрічкове представлення моделі та створити об’єкти моделі
get_system	Отримати систему рівнянь при даній фіксації змінних
get_forsythes_polynomial	Отримати поліноми Форсайта для заданої регресії
get_var_count	Отримати кількість змінних з стрічкового представлення моделі
mult_polyfit_active	Алгоритм багатовимірної регресії
get_experiment_plan	Згенерувати план експериментів
terms_from_str	Привести вхідні дані до формату, зрозумілого алгоритму
generate_report	Згенерувати звіт, який містить результати регресійного аналізу
is_singular	Перевірка на виродженість СЛАР

Для коректної роботи програми файл із вхідними даними задачі має відповідати визначеному форматові. При створенні плану експериментів файл має відповідати формату вкладеного JSON-Array, що містить всі передбачувані члени поліному. Формат кожного елементу являє собою JSON-Array розмірністю що відповідає кількості змінних n_vars та містить ступінь змінної відповідно до індексу у масиві.

Цей формат наведено на рисунку 4.5.

```
[
[t1_x1_power, t1_x2_power, … t1_xn_power],
[t2_x1_power, t2_x2_power, … t2_xn_power],
…
[tm_x1_power, tm_x2_power, … tm_xn_power]
]
```

Рисунок 4.5 – Формат вхідних даних

При знаходженні лінії регресії, формат вхідних даний є складнішим JSON-об’єктом, та містить масив із даних про експерименти, де кожен елемент містить наступну інформацію:

- boolean-масив що позначає фіксовані змінні *fixations*;
- значення зафіксованих зміних *fixed_values*;
- масив вхідних змінних *x_data;*
- масив вихідних змінних *y_data;*

Приклад даних за цим форматом наведено на рисунку 4.6.

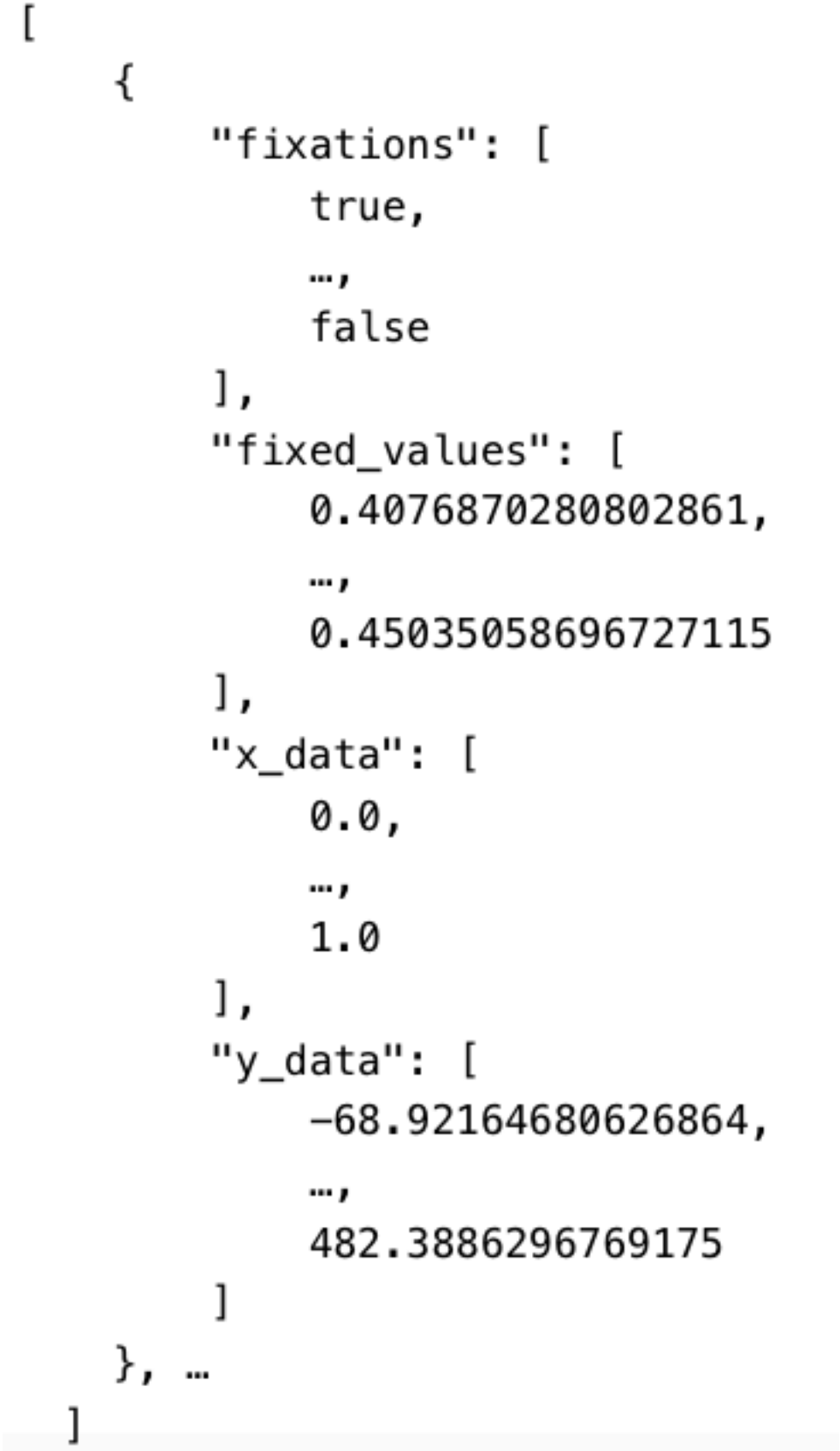

Рисунок 4.6 – Формат вхідних даних

На рисунку 4.6 наведено приклад вихідного файлу, який створює програма при завантаженні отриманого одним з алгоритмів розв’язку.

```
Результати розв'язування задачі 100_5x200.txt алгоритмом AIB.

Сумарна важливість: 2839
Розподіл завдань:

Учасник №1: [ завдання №47, час початку: 0, час виконання: 27],
[ завдання №63, час початку: 27, час виконання: 19], [ завдання №80, час
початку: 46, час виконання: 5], [ завдання №100, час початку: 51, час
виконання: 4], [ завдання №109, час початку: 55, час виконання: 5],
[ завдання №132, час початку: 60, час виконання: 33], [ завдання №161,
час початку: 93, час виконання: 3], [ завдання №31, час початку: 96, час
виконання: 1], [ завдання №46, час початку: 97, час виконання: 1],
[ завдання №95, час початку: 98, час виконання: 1], [ завдання №192, час
початку: 99, час виконання: 1].

Учасник №2: [ завдання №18, час початку: 0, час виконання: 19],
[ завдання №50, час початку: 19, час виконання: 27], [ завдання №55, час
початку: 46, час виконання: 14], [ завдання №106, час початку: 60, час
виконання: 8], [ завдання №120, час початку: 68, час виконання: 20],
[ завдання №150, час початку: 88, час виконання: 12].

Учасник №3: [ завдання №17, час початку: 0, час виконання: 27],
[ завдання №22, час початку: 27, час виконання: 23], [ завдання №36, час
початку: 50, час виконання: 2], [ завдання №41, час початку: 52, час
виконання: 27], [ завдання №91, час початку: 79, час виконання: 11],
[ завдання №144, час початку: 90, час виконання: 6], [ завдання №39, час
початку: 96, час виконання: 1], [ завдання №105, час початку: 97, час
виконання: 1], [ завдання №35, час початку: 98, час виконання: 1].

Учасник №4: [ завдання №152, час початку: 0, час виконання: 46],
[ завдання №128, час початку: 46, час виконання: 1], [ завдання №89, час
початку: 47, час виконання: 1], [ завдання №6, час початку: 48, час
виконання: 19], [ завдання №122, час початку: 67, час виконання: 16],
[ завдання №103, час початку: 83, час виконання: 14].

Учасник №5: [ завдання №42, час початку: 0, час виконання: 3],
[ завдання №12, час початку: 3, час виконання: 10], [ завдання №163, час
початку: 13, час виконання: 3], [ завдання №72, час початку: 16, час
виконання: 22], [ завдання №69, час початку: 38, час виконання: 1],
[ завдання №189, час початку: 39, час виконання: 16], [ завдання №195,
час початку: 55, час виконання: 32].
```

Рисунок 4.6 – Приклад вихідного файлу з розв’язком, створеного програмою

Як видно з рисунку 4.6, програма виводить сумарну важливість обраних завдань та розклад для кожного учасника.

У додатку 3 наведено схему структурну послідовності роботи програми, , у додатку YY – загальну схему структурну класів розробленого програмного забезпечення, у додатку 5 – схему структурну пакетів.

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ АЛГОРИТМІВ

ВИСНОВКИ

В статье наведено конструктивный метод відновлення многомерной поліномиальной регресії, представленной надлишковим описом,звикористанням ограниченного активного експерименту. Показано, що при использовании модифицированного алгоритму оцінки коефіцієнтівнабагато ближе к реальным данным чим в главе 6 [1].

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Павлов А.А., Калашник В.В., Коваленко Д.А. Построение багатовимірної поліномиальной регресії. Регресія при даних з повторюваними аргументуми // Вісник НТУУ “КПІ”. Серія «Інформатика, управління та обчислювальна техніка». – К.: “ВЕК+”, 2015. – No63. – 4 с.

2. Чорний ящик [Електронний ресурс] // Режим доступу: https://uk.wikipedia.org/wiki/Чорний_ящик

3. Планування експерименту [Електронний ресурс] // Режим доступу: https://uk.wikipedia.org/wiki/Планування_експерименту

4. Худсон Д. Статистика для физиков: Лекции по теории вероятностей і елементарной статистике / Д. Худсон. – 2-е изд., доп. – М.: Мир, 1970. – 296 с.

5. Multivariate Polynomial Regression [Електронний ресурс] // Режим доступу: https://github.com/ahmetcecen/MultiPolyRegress-MatlabCentral

6. Multivariate Polynomial Regression [Електронний ресурс] // Режим доступу: http://www.mathworks.com/matlabcentral/fileexchange/34918-multivariate-polynomial-regression/all_files

7. Online Multiple Polynomial Regression [Електронний ресурс] // Режим доступу: http://www.xuru.org/rt/mpr.asp

8. ECMA-404 The JSON Data Interchange Standard [Електронний ресурс] // Режим доступу: http://www.json.org

9. JSON [Електронний ресурс] // Режим доступу: https://en.wikipedia.org/wiki/JSON

10. Orthogonal polynomials [Електронний ресурс] // Режим доступу: https://en.wikipedia.org/wiki/Orthogonal_polynomials

11. Згуровский М. З. Принятие решений в сетевых системах з ограниченными ресурсами [Текст] : [монография] / М. З. Згуровский, А. А. Павлов ; Нац. техн. ун-т "Киев. политехн. ин-т". - К. : Наукова думка, 2010. - 575 с. : рис., табл. - Бібліогр.: с. 560-569.